

"十三五"普通高等教育本科规划教材

21世纪全国高等院校物流专业创新型应用人才培养规划教材

LOGISTICS

（第2版）

物流运筹学



郝　海　熊德国◎主编







“十三五”普通高等教育本科规划教材
21 世纪全国高等院校物流专业创新型应用人才培养规划教材

物 流 运 筹 学
(第 2 版)

主　编　郝　海　熊德国
副主编　王永建　贲立欣　王全文

北京大学出版社版权所有
禁止转载

北京大学出版社
PEKING UNIVERSITY PRESS

内 容 简 介

本书系统介绍了运筹学中的经典内容，重点讲解了应用最为广泛的线性规划、图与网络、网络计划、决策论、动态规划、排队论、存储论等定量分析方法。本书以问题为导向，帮助学生理清管理问题的结构，量化问题中的关系；注重理论和应用的平衡，既深入浅出地介绍运筹学的思想，又通过理论联系实际来阐述管理问题的数学建模方法，对于严格的数学推导过程则是点到即止。同时，在应用中插入了大量物流活动的实例，使学生能够了解运筹技术实际的、确切的应用背景，并设计了应用软件解决管理问题的实际操作训练。

本书条理清晰、通俗易懂，适合作为物流工程、物流管理、管理工程、交通运输、信息管理等专业的本科或研究生教材，也可作为物流咨询公司、物流企业的物流从业者和其他企业物流部门管理人员的参考书。

图书在版编目(CIP)数据

物流运筹学/郝海，熊德国主编. —2版. —北京：北京大学出版社，2017.3
（21世纪全国高等院校物流专业创新型应用人才培养规划教材）
ISBN 978-7-301-28110-9

Ⅰ. ①物…　Ⅱ. ①郝…②熊…　Ⅲ. ①物流—运筹学—高等学校—教材　Ⅳ. ①F252

中国版本图书馆CIP数据核字（2017）第029604号

书　　名　物流运筹学（第2版）
WULIU YUNCHOUXUE
著作责任者　郝　海　熊德国　主编
责任编辑　刘　丽
标准书号　ISBN 978-7-301-28110-9
出版发行　北京大学出版社
地　　址　北京市海淀区成府路205号　100871
网　　址　http://www.pup.cn　新浪微博：@北京大学出版社
电子信箱　pup_6@163.com
电　　话　邮购部62752015　发行部62750672　编辑部62750667
印 刷 者
经 销 者　新华书店
787毫米×1092毫米　16开本　20.25印张　462千字
2010年8月第1版
2017年3月第2版　　2018年1月第2次印刷
定　　价　45.00元

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究
举报电话：010-62752024　电子信箱：fd@pup.pku.edu.cn
图书如有印装质量问题，请与出版部联系，电话：010-62756370

北京大学出版社版权所有 禁止转载

21世纪全国高等院校物流专业创新型应用人才培养规划教材

编写指导委员会

(按姓名拼音顺序)

主任委员	齐二石			
副主任委员	白世贞	董千里	黄福华	李荷华
	王道平	王槐林	魏国辰	徐　琪
委　　员	曹翠珍	柴庆春	丁小龙	甘卫华
	郝　海	阚功俭	李传荣	李学工
	李於洪	林丽华	柳雨霁	马建华
	孟祥茹	倪跃峰	乔志强	汪传雷
	王海刚	王汉新	王　侃	吴　健
	易伟义	于　英	张　军	张　浩
	张　潜	张旭辉	赵丽君	周晓晔

北京大学出版社版权所有
禁止转载

北京大学出版社版权所有
禁止转载

丛 书 总 序

物流业是商品经济和社会生产力发展到较高水平的产物，它是融合运输业、仓储业、货代业和信息业等的复合型服务产业，是国民经济的重要组成部分，涉及领域广，吸纳就业人数多，促进生产、拉动消费作用大，在促进产业结构调整、转变经济发展方式和增强国民经济竞争力等方面发挥着非常重要的作用。

随着我国经济的高速发展，物流专业在我国的发展很快，社会对物流专业人才需求逐年递增，尤其是对有一定理论基础、实践能力强的物流技术及管理人才的需求更加迫切。同时随着我国教学改革的不断深入以及毕业生就业市场的不断变化，以就业市场为导向，培养具备职业化特征的创新型应用人才已成为大多数高等院校物流专业的教学目标，从而对物流专业的课程体系以及教材建设都提出了新的要求。

为适应我国当前物流专业教育教学改革和教材建设的迫切需要，北京大学出版社联合全国多所高校教师共同合作编写出版了本套“21 世纪全国高等院校物流专业创新型应用人才培养规划教材”。其宗旨是：立足现代物流业发展和相关从业人员的现实需要，强调理论与实践的有机结合，从“创新”和“应用”两个层面切入进行编写，力求涵盖现代物流专业研究和应用的主要领域，希望以此推进物流专业的理论发展和学科体系建设，并有助于提高我国物流业从业人员的专业素养和理论功底。

本系列教材按照物流专业规范、培养方案以及课程教学大纲的要求，合理定位，由长期在教学第一线从事教学工作的教师编写而成。教材立足于物流学科发展的需要，深入分析了物流专业学生现状及存在的问题，尝试探索了物流专业学生综合素质培养的途径，着重体现了“新思维、新理念、新能力”三个方面的特色。

1. 新思维

(1) 编写体例新颖。借鉴优秀教材特别是国外精品教材的写作思路、写作方法，图文并茂、清新活泼。

(2) 教学内容更新。充分展示了最新最近的知识以及教学改革成果，并且将未来的发展趋势和前沿资料以阅读材料的方式介绍给学生。

(3) 知识体系实用有效。着眼于学生就业所需的专业知识和操作技能，着重讲解应用型人才培养所需的内容和关键点，与就业市场结合，与时俱进，让学生学而有用，学而能用。

2. 新理念

(1) 以学生为本。站在学生的角度思考问题，考虑学生学习的动力，强调锻炼学生的思维能力以及运用知识解决问题的能力。

(2) 注重拓展学生的知识面。让学生能在学习到必要知识点的同时也对其他相关知识有所了解。

(3) 注重融入人文知识。将人文知识融入理论讲解，提高学生的人文素养。

北京大学出版社版权所有
禁止转载

3. 新能力

(1) 理论讲解简单实用。理论讲解简单化，注重讲解理论的来源、出处以及用处，不做过多的推导与介绍。

(2) 案例式教学。有机融入了最新的实例以及操作性较强的案例，并对案例进行有效的分析，着重培养学生的职业意识和职业能力。

(3) 重视实践环节。强化实际操作训练，加深学生对理论知识的理解。习题设计多样化，题型丰富，具备启发性，全方位考查学生对知识的掌握程度。

我们要感谢参加本系列教材编写和审稿的各位老师，他们为本系列教材的出版付出了大量卓有成效的辛勤劳动。由于编写时间紧、相互协调难度大等原因，本系列教材肯定还存在不足之处。我们相信，在各位老师的关心和帮助下，本系列教材一定能不断地改进和完善，并在我国物流专业的教学改革和课程体系建设中起到应有的促进作用。

齐二石

2009 年 10 月

齐二石　本系列教材编写指导委员会主任，博士、教授、博士生导师。天津大学管理学院院长，国务院学位委员会学科评议组成员，第五届国家 863/CIMS 主题专家，科技部信息化科技工程总体专家，中国机械工程学会工业工程分会理事长，教育部管理科学与工程教学指导委员会主任委员，是最早将物流概念引入中国和研究物流的专家之一。

北京大学出版社版权所有
禁止转载

北京大学出版社版权所有
禁止转载

目　　录

北京大学出版社版权所有
禁止转载

北京大学出版社版权所有 禁止转载

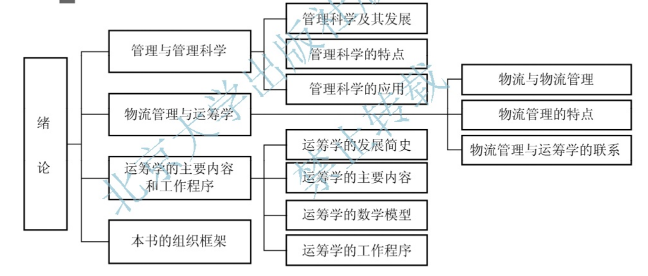
北京大学出版社版权所有
禁止转载

北京大学出版社版权所有
禁止转载

第1章 绪论



【本章知识架构】



北京大学出版社版权所有 禁止转载

【本章教学目标与要求】

- 了解管理科学的内涵、特点和应用。
- 理解物流与物流管理内涵，掌握物流管理的特点，了解物流管理与运筹学的关系。
- 了解运筹学发展历史，熟悉运筹学的研究内容，掌握运筹学的工作程序。
- 了解本书的组织框架。

导入案例

现 代 物 流

物流水平是一个国家经济发展水平和现代文明的标志。现代物流贯穿于生产、流通、消费等社会再生产的各个环节之中，物流是经济运行的润滑剂和加速剂，由于现代物流能够节约流通成本，提供增值服务，创造竞争优势，在国际上普遍将其称为“降低成本的最后边界”。

物流早在我国“十五”计划中就被列为大力发展的新型服务业之一，也是近期国务院发布的“十大产业振兴规划”中的产业之一。现代物流在我国起步较晚，潜在需求巨大，2015 年社会物流总费用与 GDP 的比率为 16.6%，而目前国外发达市场经济国家这一比率大约为 10%，这表明我国物流运作管理水平是极为粗放的，社会化、专业化水平低，经济增长付出的物流成本太高。只有不断提高我国物流的管理效率，才能加速生产与流通领域的集约化程度，改善粗放型的经营和管理方式，促进我国经济持续、快速、健康发展。

经济全球化进程的加速，企业面临着越来越复杂的市场环境，顾客的消费需求日益多元化和个性化，资源在全球范围内最佳配置。企业间的竞争不断加剧，要求企业实现对市场的快速反应，提高其顾客服务水平；要求各个领域精心地处理各种管理问题，重视管理的科学内涵，不断提升其管理水平。

北京大学出版社版权所有 禁止转载

1.1 管理与管理科学

管理就是管理者运用各种资源达成既定目标的过程，在这一过程中，只有管理者具备一定的科学素养，才能更有效地运用有限的资源达到更高水平的目标。管理实践必须依靠科学的管理，掌握系统化的管理知识，管理者才有可能对管理中存在的问题找到可行的正确的解决方案。

1.1.1 管理科学及其发展

管理科学是研究管理理论、方法和管理实践活动的一般规律的科学。管理科学的初创阶段始于 19 世纪末 20 世纪初，由美国工程师费雷德里克·泰罗创建出“标准劳动方法”和劳动定额，被称为“泰罗制”，他于 1911 年发表了代表作《科学管理原理》；与“科学管理理论”同期问世的还有法约尔的“管理过程理论”和韦伯的“行政组织理论”，这三种理论统称为“古典管理理论”。管理科学的第二个里程碑是“行为科学理论”，诞生于 20 世纪 20 年代，创始人是美国哈佛大学教授乔治·奥尔顿·梅奥和费里茨·罗特利斯伯格等。

到了 20 世纪 50 年代管理科学的基本方法已经形成，美国于 1953 年成立管理科学学会，出版会刊《管理科学》。20 世纪 60 年代后，管理科学又运用行为科学的原理扩大到人事的组织和决策。管理科学在广泛应用过程中，同许多社会科学学科和自然科学学科交叉、渗透，产生了种种管理学分支，例如：管理社会学、行政管理学、军事管理学、教育管理学、卫生管理学、技术管理学、城市管理学、国民经济管理学等。20 世纪 80 年代管理科学已涉及战略规划和战略决策，以进一步优化组织和管理，提高效益。当今管理科学已经扩展

到各个领域，形成了内容广泛、门类齐全的独立学科体系，管理科学已经成为同社会科学、自然科学并列的第三类科学。

1.1.2　管理科学的特点

在经济管理中，管理的自然属性表现为科学合理地组织生产力要素，处理和解决经济活动中物与物、人与物之间的技术联系，如生产中的配料问题、生产力布局、规划，以及机器设备的技术性能对操作者的技术水平和熟练程序的要求等，都体现自然规律和技术规律的要求，不受社会的经济基础和上层建筑的影响；而管理的社会属性则表现为调和与完善生产关系，合理地调整人与人之间的经济利益关系，如分配体制、管理体制等，都由社会、经济规律支配。在现代经济的发展中，科学管理起着越来越重要的作用，科学管理带来了直接的经济效益，特别是在物质资源稀缺的情况下，管理的作用显得尤为重要。

管理科学具有以下特点。

(1) 现代管理理论以系统论、信息论、控制论为理论基础，以“系统理论”“决策理论”“管理科学”等学派为代表。

(2) 力求减少决策中的个人主观成分。依靠建立一套决策程序和数学模型来增加决策的科学性。他们将众多方案中的各种变数或因素加以数量化，利用数学工具建立数量模型，研究各变数和因素之间的相互关系，寻求一个用数量表示的最优解决方案。

(3) 各种可行的解决方案均是以经济效益作为评价的依据。例如成本、总收入和投资利润率等。

(4) 广泛地使用计算机。现代企业管理中影响某一事务的因素错综复杂，建立模型后，计算任务极为繁重，依靠传统的计算方法获得结果往往需要大量时间，致使计算结果无法用于企业管理。计算机的出现大大提高了运算的速度，使数学模型应用于企业和组织成为可能。

管理科学的发展过程也反映了管理科学研究重心的转移，例如：古典管理阶段侧重于物、财及管理组织过程的管理，其研究方法是以工业工程方法及经济学方法为基础的；人际关系学说、行为科学则是建立在心理学、社会学和人类学等学科方法论基础之上的；现代管理阶段则更是吸取了数学、计算机科学技术及其他自然科学、统计学、系统论、文化学、法学等诸多学科的方法论。

北京大学出版社版权所有 禁止转载

1.1.3　管理科学的应用

管理科学对许多国家的生产力的提高作出了重要的贡献，改善了许多组织的生产效率，在实际应用中产生了巨大的效益，每年为管理科学实践者颁发的弗兰茨·厄德曼(Franz Edelman)奖被授予全世界管理科学的最佳应用项目的管理者。

弗兰茨·厄德曼奖是由世界著名的管理科学家 Franz Edelman 于 1971 年创立的，弗兰茨·厄德曼奖的评选范围包括管理科学理论和方法的创新、应用工作对企业创造的直接经济效益，以及对社会和人类生活所作的积极贡献，全世界各个国家的科学家和企业家都可以申请这一奖项。在 2002 年我国于刚教授领导的项目小组由于对实时决策支持系统理论和方法的创立和推进，为大陆航空公司等民用航空企业创造了实实在在的经济效益，成为本年度大奖的获得者，其他年度获奖项目见表 1-1。

表 1-1　获得弗兰茨·厄德曼奖的管理科学应用项目

年　　度	组　　织	项目内容	每年节支/百万美元
2016	UPS	借助于 UPS 道路综合优化和导航系统，自动规划 5.5 万司机复杂的配送路线	300～400
2012	TNT 快递	着力于全球运输网络的优化，包括选址计划、最优卡车路线、车队管理和员工作息安排等	50
2009	HP 公司	开发出收入覆盖优化(RCO)软件算法和复杂投资回报计算器用来进行产品多样化管理，以满足不同顾客的需求	500
2008	荷兰铁路	重新修订的火车时刻表提高了原本已极其繁忙的铁路联网的运输能力，而同时又无须建设更多的铁道	60
2006	华纳罗宾逊航空物流中心	维护与检修 C-5 运输机，维修时间降低 33%	49.8

资料来源：http://www.informs.org/.

北京大学出版社版权所有 禁止转载

1.2　物流管理与运筹学

现代物流是涉及运输、仓储、货代、联运、制造、贸易、信息等相关行业的新兴综合性产业，物流活动的内容纷繁复杂，这就要求物流活动中的相关企业科学合理地安排物流活动的各项内容。为降低物流成本，提高运作效率，不可避免地运用到各种定量管理的科学技术和方法，运筹学是其中非常重要的一门应用学科。

1.2.1　物流与物流管理

物流管理是近一二十年以来兴起的一门新学科，它是管理科学新的重要分支。物流是从企业传统的生产和销售活动中分离出来的，成为独立的研究领域和学科范围。物流科学的诞生使得原来在经济活动中处于潜隐状态的物流系统显现出来，它揭示了物流活动的各个环节的内在联系，它的发展和日臻完善，是现代企业在市场竞争中制胜的法宝。

1. 物流与物流管理

物流(Logistics)是指对商品、服务及相关信息在供应地到接收地之间，有效率和有效益地正向和反向移动与储存，进行的计划、执行和控制的过程，主要包括运输、储存、装卸、搬运、包装、流通加工、配送、信息处理等基本功能。现代物流成为推动流通现代化的一种重要方式。

物流管理(Logistics Management)是指在社会再生产过程中，根据物质资料实体流动的规律，应用管理的基本原理和科学方法，对物流活动进行计划、组织、指挥、协调、控制和监督，使各项物流活动实现最佳的协调与配合，以降低物流成本，提高物流效率和经济效益。现代物流管理是建立在系统论、信息论和控制论的理论基础上的。

实施物流管理的目的就是要在尽可能最低的总成本条件下实现既定的客户服务水平，即寻求服务优势和成本优势的一种动态平衡。物流管理要解决的基本问题，简单地说，就是把合适的产品以合适的数量和合适的价格在合适的时间与合适的地点提供给客户。

2. 物流管理发展的动因

20 世纪 90 年代以来，企业的外界市场环境发生了很大的变化，如客户服务激增、服务时间成为管理的焦点，迫使企业越来越重视物流管理。由于市场竞争的日趋激烈，人们发现在生产过程中降低成本获得竞争优势的空间已经非常有限了，这样就迫使企业开始把寻求成本优势和服务差别化优势的目光转向生产前后延伸的物流领域，并开始在该领域实行全程质量管理。可以说，从世界范围看，在当今企业面临的诸多战略问题中，最具挑战的领域之一就是物流管理。

现代物流管理科学的不断发展和完善，有以下几个在技术、经济和政策上的原因。

(1) 产品越来越趋向多品种、小批量生产，产品的多样性和小批量化决定了配送的复杂化。

(2) 制造和销售中 JIT、快速响应、连续补货等技术的采用，引进了以时间为基本条件的物流服务，要求人们做到在库存、运输和生产、销售之间进行严格控制与协调，以使存货量能够降低到最低限度，改变传统上依赖于安全库存的物流战略。

(3) 计算机技术和信息技术的发展，有力地推动了现代物流的发展。物流管理高度依赖于对大量的数据、信息的分析和处理。计算机技术与网络技术的发展和商业化为物流系统提供了分析问题、处理事务、进行评估和决策的支持处理平台。条形码技术、电子数据交换(Electronic Data Interchange，EDI)、卫星通信技术等信息技术在物流作业中被广泛应用，大大提高了物流服务的水平和物流过程的透明度。

(4) 发达国家在运输领域广泛采取的“放松管制”政策，使运输市场的竞争空前激烈，以客户服务为中心的物流创新层出不穷。

北京大学出版社版权所有 禁止转载

3. 物流管理的主要内容和研究领域

物流管理包括下述内容：①对物流活动诸要素的管理，包括运输、储存等环节的管理；②对物流系统诸要素的管理，即对其中人、财、物、设备、方法和信息六大要素的管理；③对物流活动中具体职能的管理，主要包括物流计划、质量、技术、经济等职能的管理等。

由于物流对象不同，物流目的不同，物流范围不同，形成了不同的物流研究领域。

1) 宏观物流

宏观物流是指社会再生产总体的物流活动，从社会再生产总体角度认识和研究的物流活动。宏观物流还可以从空间范畴来理解，在很大空间范畴的物流活动，往往具有宏观性。

2) 微观物流

消费者、生产者企业所从事的实际的、具体的物流活动属于微观物流。在整个物流活动中的一个局部、一个环节的具体物流活动也属于微观物流。在一个小地域空间发生的具体的物流活动也属于微观物流。

3) 企业物流

从企业角度上研究与之有关的物流活动，是具体的、微观的物流活动的典型领域。

4) 供应链物流

供应链物流是以物流活动为核心，协调供应领域的生产和进货计划、销售领域的客户服务和订货处理业务，以及财务领域的库存控制等活动，包括了对涉及采购、外包、转化

等过程的全部计划和管理活动，以及全部物流管理活动。

5) 国际物流

国际物流是现代物流系统发展很快、规模很大的一个物流领域，国际物流是伴随和支撑国际经济交往、贸易活动和其他国际交流所产生的物流活动。

6) 区域物流

相对于国际物流而言，一个国家范围内的物流，一个城市的物流，一个经济区域的物流都处于同一法律、规章、制度之下，都受相同文化及社会因素影响，都处于基本相同的科技水平和装备水平之中。

1.2.2 物流管理的特点

现代物流是与现代经济社会紧密联系在一起的，体现了现代企业经营和社会经济发展的需要。在现代物流管理和运作中，广泛采用代表着当今生产力发展水平的管理技术、工程技术，以及信息技术等。现代物流的本质是服务，它本身并不创造商品的形质效用，而是产生空间效用、时间效用和质量效用。物流服务是企业为了满足客户(包括内部和外部客户)的物流需求，开展一系列物流活动的结果。物流管理具有以下特点。

1. 物流管理的系统化

物流管理强调运用系统方法解决问题，物流管理既重视效率更注重效果。现代物流通常被认为是由运输、储存、包装、装卸、流通加工、配送和信息处理诸环节构成，各环节原本都有各自的功能、利益和观念。物流管理就是利用现代管理方法及其技术，使各个环节共享总体信息，把所有环节作为一体化的系统来组织和管理，以使系统能够在尽可能低的总成本条件下，提供有竞争优势的客户服务。

北京大学出版社版权所有
禁止转载

2. 物流管理的信息化

物流管理信息化是指建立高效畅通的物流信息系统，实现从物流决策、业务流程、客户服务的全程信息化，从而对物流活动进行科学管理。物流系统只有具有良好的信息处理和传输系统，才能快速、准确地获取销售反馈信息和配送货物跟踪信息，从而大大提高物流企业的服务水平。在物流信息系统的建设中，一方面要重视新的信息技术的应用，提高信息技术的水平；另一方面也要重视物流信息系统和物流管理的互动，既要根据自己的物流管理流程来选择适合的物流信息系统，也要通过物流信息系统来优化和再造自己的物流管理流程。

3. 物流管理的智能化

智能化就是利用条形码、射频识别技术、传感器、全球定位系统等先进的物联网技术通过信息处理和网络通信技术平台广泛应用于物流业运输、仓储、配送、包装、装卸等基本活动环节，实现货物运输过程的自动化运作和高效率优化管理，降低成本，减少自然资源和社会资源消耗。物联网和云计算为物流业将传统物流技术与智能化系统运作管理相结合提供了一个很好的渠道，智能物流在功能上要实现 6 个“正确”，即正确的货物、正确的数量、正确的地点、正确的质量、正确的时间、正确的价格；在技术上要实现：物品识别、地点跟踪、物品溯源、物品监控、实时响应。

1.2.3　物流管理与运筹学的联系

物流管理是管理科学的重要分支，对物流活动的管理必然要继承和发展管理科学的分析方法。现代物流所呈现的复杂性不是简单算术所能解决的，以计算机为手段的运筹技术正渐渐成为支撑现代物流管理的有效工具。现代物流的发展离不开运筹学的技术支持，运筹学的应用将会使物流管理更加高效；现代物流也为运筹技术的发展与完善提供了现实背景，为运筹技术展现了广阔的应用空间。

物流管理与运筹学具有天然的联系，从两者产生的时间看，都是在第二次世界大战时期为军事所重视而利用发展起来的，如图 1.1 所示。



图 1.1　运筹学一些主要分支产生的时间线

从功能上来说，运筹学是用来解决最优资源配置，而物流管理的主要目标也正是追求一种快速、及时、节约、库存合理的物流服务，这一点正好不谋而合。两者自始至终都密切地联系在一起，互相渗透和交叉发展，虽说后来一段时间相对于运筹学的发展，物流管理相对滞后，但随着全球经济的不断发展，物流管理活动中运筹学的应用不断扩大，运筹学的作用渐渐凸现。

北京大学出版社版权所有　禁止转载

1.3　运筹学的主要内容和工作程序

运筹学是管理科学的重要理论基础和应用手段，运筹学主要是将生产、管理等活动中出现的一些带有普遍性的问题加以提炼，然后利用数学方法进行解决。

1.3.1　运筹学的发展简史

运筹学的英文通用名称为 Operation Research，简写为 OR，原意为运作研究或作战研究，中国科学家把它译成“运筹学”，巧妙地借用了《史记·高祖本纪》中“运筹帷幄之中，决胜千里之外”的典故。运筹学的思想由来已久，我国历史上在军事和科学技术方面对运筹思想的运用是世界闻名的。公元前 6 世纪春秋时期著名的《孙子兵法》中处处体现了军事运筹的思想；战国时期的“田忌赛马”故事是对策论的典型范例；除军事方面，在我国古代农业、运输、工程技术等方面也有大量体现运筹学思想的实例，如北魏时期科学家贾思勰的《齐民要术》就是一部体现运筹学思想、合理策划农事的宝贵文献；古代的粮食和物资的调运，都市的规划建设，如“一举而三役济”的“丁渭造宫”，水利方面如四川都江堰工程，无不渗透了运筹思想的运用。

运筹学是 20 世纪 40 年代开始形成的一门应用数学学科，起源于第二次世界大战期间英、美等国的军事运筹小组。在第二次世界大战初期，英美两国的军事部门迫切需要研究如何将非常有限的人力和物力分配到各项军事活动中，以达到最好的作战效果。1935 年，英国为了对付德国越来越严重的空中力量威胁，英国防空科学调查委员会组织了一

些科学家专门研究如何使用雷达来进行对空作战的问题，有效地遏制了德国空军的进攻。作战研究部主任 A. P. Rowe 把他们从事的工作称为 Operational Research，美国则称为 Operations Research。在第二次世界大战期间，运筹学成功地解决了许多重要作战问题，比较著名的有大西洋海战、不列颠空战等，均显示了运筹学的巨大威力，运筹学在此后得以迅速发展。

第二次世界大战结束后，那些从事作战研究的人员纷纷转入工业生产部门和商业部门。由于组织内部与日俱增的复杂性和专门化所产生的问题，使人们认识到这些问题本质上与战争中曾面临的问题极为相似，只是具有不同的现实环境而已，运筹学于是进入工商企业和其他部门。1950 年，英国的伯明翰大学正式开设了运筹学课程，同年第一本运筹学杂志《运筹学季刊》(*O.R. Quarterly*)在英国创刊；1951 年，美国的莫尔斯(P.M. Morse)和金博尔(G.E. Kimball)合著的《运筹学方法》正式出版；1952 年，美国运筹协会成立，并于同年出版了运筹学杂志(*Journal of ORSA*)，所有这些标志着该门学科基本形成，随后这门学科的理论体系也在不断完善。

20 世纪 50 年代后期，我国著名科学家钱学森、华罗庚、许国志等将运筹学引入中国，并结合我国的特点在国内推广应用，著名的“打麦场的选址问题”和“中国邮递员问题”就是在那个时期提出的。华罗庚院士自 1965 年起与他的学生一起走出研究所，用十年的时间在全国推广“优选法”和“统筹法”，对中国运筹学的研究和应用起到了巨大的推动作用。计算机的问世、高速化发展与广泛普及，使得各行业从业人员能够运用这些先进的方法理论解决大量的大规模问题，从而促进了运筹学的发展并使其应用范围日益扩大。

经过半个多世纪的发展，运筹学的内容日趋成熟，逐渐形成了其理论与方法的基本框架，运筹学已经成为各行各业进行管理活动的一个基本工具。

北京大学出版社版权所有 禁止转载

1.3.2 运筹学的主要内容

运筹学作为一门实践应用的科学，已被广泛应用于工业、农业、商业、交通运输业、民政事业、军事决策等组织，解决由多种因素影响的复杂大型问题。目前，在物流领域中的应用也相当普遍，并且解决了许多实际问题，取得了很好的效果。下面总结的是运筹学在物流领域中应用较多的若干分支。

1. 数学规划论

数学规划论主要包括线性规划、非线性规划、整数规划、目标规划和动态规划。研究内容与生产活动中有限资源的分配有关，在组织生产的经营管理活动中，具有极为重要的地位和作用。它们解决的问题都有一个共同特点，即在给定的条件下，按照某一衡量指标来寻找最优方案，求解约束条件下目标函数的极值(极大值或极小值)问题。具体来讲，线性规划可解决物资调运、配送和人员分派等问题；整数规划可以求解完成工作所需的人数、机器设备台数和厂、库的选址等；动态规划可用来解决诸如最优路径、资源分配、生产调度、库存控制、设备更新等问题。

2. 图(网络)论

图论广泛应用于解决工程系统和管理问题，将复杂的问题用图与网络进行描述简化后再求解。图与网络理论有很强的构模能力，描述问题直观，模型易于计算实现，很方便地将一些复杂的问题分解或转化为可能求解的子问题。图与网络在物流中的应用也很显著，

其中最明显的是运输问题、物流网点间的物资调运和车辆调度时运输路线的选择、配送中心的送货、逆向物流中产品的回收、工程的最小周期和关键工序等，运用了图论中的最小生成树、最短路径、最大流、最小费用等知识，求得运输所需时间最少或路线最短或费用最省的路线。另外，工厂、仓库、配送中心等物流设施的选址问题，物流网点内部工种、任务、人员的指派问题，设备更新问题，也可运用图论的知识辅助决策者进行最优的安排。

3. 决策论

决策普遍存在于人类的各种活动之中，物流中的决策就是在占有充分资料的基础上，根据物流系统的客观环境，借助科学的数学分析、实验仿真或经验判断，在已提出的若干物流系统方案中，选择一个合理、满意方案的决断行为。如制订投资计划、生产计划、物资调运计划、选择自建仓库或租赁公共仓库、自购车辆或租赁车辆等。物流决策多种多样，复杂或简单，按照不同的标准可划分很多种类型，其中按决策问题目标的多少可分为单目标决策和多目标决策。单目标决策目标单一，相对简单，求解方法也很多，如线性规划、非线性规划、动态规划等；多目标决策相对而言复杂得多，如要开发一块土地建设物流中心，既要考虑设施的配套性、先进性，还要考虑投资大小问题等，这些目标有时相互冲突，这时就要综合考虑。

4. 排队论

排队论也称随机服务理论，主要研究各种系统的排队队长、等待时间和服务等参数，解决系统服务设施和服务水平之间的平衡问题，以较低的投入求得更好的服务。排队现象在现实生活中普遍存在，物流领域中也司空见惯，如工厂生产线上的产品等待加工，在制品、产成品排队等待出入库作业，运输场站车辆进出站的排队，客服中心顾客电话排队等待服务，商店顾客排队付款等。

北京大学出版社版权所有
禁止转载

5. 存储论

存储论又称库存论，主要是研究物资库存策略的理论，即确定物资库存量、补货频率和一次补货量。合理的库存是生产和生活顺利进行的必要保障，可以减少资金的占用，减少费用支出和不必要的周转环节，缩短物资流通周期，加速再生产的过程等。在物流领域中的各节点：工厂、港口、配送中心、物流中心、仓库、零售店等都或多或少地保有库存，为了实现物流活动总成本最小或利益最大化，大多数人们都运用了存储理论的相关知识，以辅助决策。

6. 对策论

对策论也称博弈论，对策即在竞争环境中作出的决策。对策论探究有多方参加的竞争活动，对策论最初是运用数学方法研究有利害冲突的双方在竞争性的活动中是否存在战胜对方的最优策略，以及如何找出这些策略等问题。在这些问题中，把双方的损耗用数量来描述，并找出双方最优策略。

常言道“商场如战场”，在市场经济条件下，物流业也充满了竞争。对策论是一种定量分析方法，可以寻找最佳的竞争策略，以便战胜对手或者减少损失。例如在一个城市内有两个配送中心经营相同的业务，为了争夺市场份额，双方都有多个策略可供选择，可以运用对策论进行分析，寻找最佳策略。例如，面对一次次的铁路大提速，公路运输公司要与

铁路系统争夺客源，有多种策略可供选择，也可用对策论研究竞争方案等。

1.3.3　运筹学的数学模型

模型是对客观世界中某一现象或事件的一种抽象表述，这种抽象就是对现象与事件起关键作用因素的提取过程。模型的形式多种多样，在管理科学中的模型多表现为数学形式，即数学模型，它是数学理论与实际问题相结合的一门科学，将管理活动问题归结为相应的数学问题，利用数学的概念、方法和理论进行深入的分析和研究，从而从定性或定量的角度来刻画实际问题，并为解决这些问题提供精确的数据或可靠的指导。

运筹学广泛地采用数学模型来分析现实的管理问题，根据研究问题的内在联系，揭示那些重要的相关变量，用变量之间函数关系表达它们之间的相互影响。运筹学模型表示一个要研究的系统，其数学模型的一般形式为

$$P=(C_1,C_2,\cdots,C_m;u_1,u_2,\cdots,u_n)$$

式中：$C_1,C_2,\cdots,C_m$ 为不受控制的系统变量(不可控变量)；$u_1,u_2,\cdots,u_n$ 为受到系统控制的变量(可控变量)；P 为对系统的性能或有效性的量度(结果变量)。

1. 模型的构成要素

一般来说，运筹学模型的最基本构成要素由三部分组成，即结果变量、决策变量和不可控变量。结果变量反映了系统目标 P 的有效性程度，它依赖于决策变量的取值；决策变量描述了决策问题中可作出选择的要素，即系统的可控变量；不可控变量是指系统环境中对决策有重要影响但不可控的因素，如图 1.2 所示。

北京大学出版社版权所有
禁止转载

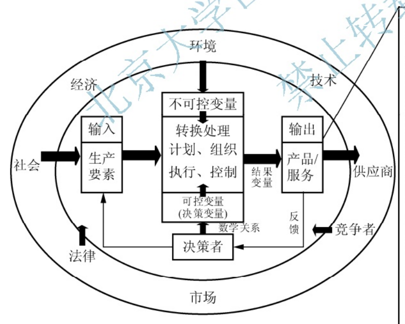


图 1.2　数学模型的系统结构

2. 模型的系统结构

模型的基本构成要素由数学关系式联系在一起，就形成了数学模型，模型的系统结构如图 1.3 所示。管理中遇到的问题不同，模型的具体表现形式也不尽相同，可能要应用各种各样的数学理论和方法，必要时还要创造新的数学理论和方法。但要注意在保证精度的

条件下尽量用简单的数学方法是建模时要遵循的一个原则，当然要求建模者对所有数学学科都精通是不可能的，但做到了解这些学科能解决哪一类问题和大体上怎样解决的方法对开阔思路是很有帮助的。此外，根据不同对象的一些相似性，借用某些学科中的数学模型，也是构建模型常使用的方法。

以最具代表性的数学规划模型为例说明数学模型的系统结构。线性规划模型中的数学关系式有两类，即目标函数与约束条件，这里以一个企业的生产计划问题为例，该企业拥有一些资源 b 来生产两类产品，市场价格是 p_1 和 p_2，追求收入最大化是企业的目标。

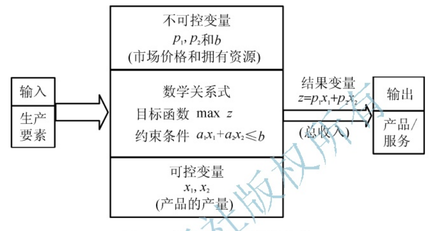


图 1.3　数学模型系统结构示例

企业生产的外部环境就是市场，它从市场中获取生产原料，并将产品送到市场中销售，市场价格是企业不能控制的，企业想提高产品价格就会导致产品卖不出去，所以市场价格是企业的不可控因素；企业在生产时拥有的资源也是不可改变的，如生产线的设计产能是企业不能控制的，短时期内想提高其生产数量也是不可能的，因为生产线安装与调试都需要一定的时间。企业的生产计划其实就是安排两种产品生产的数量，设为 x_1、x_2，这两个变量就是可控变量，企业管理者可以在前述环境下决定其生产数量，因此这两个变量称为决策变量，企业追求的是收入多少，结果变量就是 $z = p_1x_1 + p_2x_2$，企业追求的目标就是 $\max z = p_1x_1 + p_2x_2$，企业生产数量受到资源的限制，即约束 $a_1x_1 + a_2x_2 \leqslant b$，如图 1.3 所示。

1.3.4　运筹学的工作程序

应用运筹学完整地去解决管理活动中问题，一般分为 6 个步骤，如图 1.4 所示。

1. 步骤 1——明确问题

运用运筹学去解决管理活动中的实际问题首先要明确“问题是什么”，以便正确地界定问题。该步骤需要与管理决策者进行沟通互动，对他提供的问题状况进行认真研究。了解管理者所面临的困境，收集相关资料，明确究竟解决什么事情？要达到什么目标？为此必须作出什么决策。这是运筹技术与管理科学的接口——入口过程。

由于实际问题往往是复杂多样的，涉及方方面面的因素，如果把涉及的所有因素都考虑到，既不可能也没必要，而且还会使问题复杂化，所以明确问题时首先要考虑问题涉及的范围，用系统的观点和方法去分析，把握考虑问题所在系统的边界和环境，了解所分析问题中可以控制的因素，以及不可控的重要影响因素等，如图 1.2 所示。

北京大学出版社版权所有 禁止转载

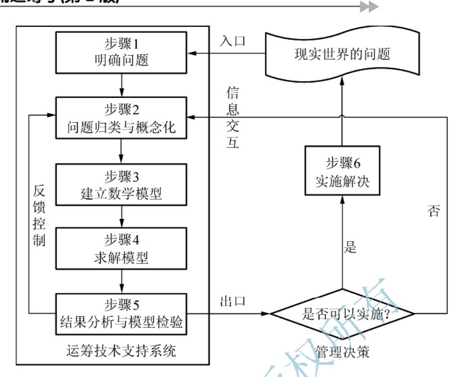


图1.4 运筹学的工作程序

北京大学出版社版权所有 禁止转载

2. 步骤2——问题归类与概念化

问题明确以后，还要对其作类型判别，即先看它是否能用运筹方法解决问题。如果是，再看它属于哪一类运筹问题，或者它适合用哪一类运筹方法解决。

3. 步骤3——建立数学模型

建立数学模型是管理科学方法的关键步骤，具体内容主要是在问题概念化的基础下进一步确定模型的构成要素以及它们之间的联系，用数学关系式表达出来，形成数学模型。

4. 步骤4——求解模型

建立的数学模型可以采用解方程、推理、图解、计算机模拟、定理证明等各种传统的和现代的数学方法对其进行求解，其中有些可以用计算机软件来做这些工作。建模的目的是寻找规律以解决实际问题，该过程称为模型求解。

管理科学的模型种类很多，解的性质及求解方法各异。“最优解”是解的主流类型，它是在模型的所有可行解中寻找出最优的一个。除了“最优解”之外，还有其他一些解的类型，如“满意解”和“描述性的解”。“满意解”是指在模型的次优解或“各行所长”的一组解中使决策者满意的解；“描述性的解”相应于描述性的模型，该类模型的目的主要是描述系统在不同条件下的状态，可用于预测和分析系统的行为特征。相对于“最优解”，这些其他类型的解也称为“非最优的解”。

5. 步骤5——结果分析与模型检验

求得模型解以后，还要对它进行分析，同时要检验模型的正确性。分析和检验的内容首先是看结果与实际情况是否相符，模型是否能够正确地反映实际问题，具体方法是将不同条件下的数据(如不同时期的数据)代入模型，看相应的解是否符合实际。其次要分析模

型中的参数发生小范围变化时对解的影响，该过程称为灵敏度分析，如果解对参数变化的反应不过于灵敏，则在实际中可能会更容易应用。

把模型在数学上分析的结果与研究的实际问题作比较以检验模型的合理性称为模型检验。如果认为模型不能很好地反映实际问题，则要重新对问题进行分析并适当修改模型。通常，一个模型要经过如此多次反复修改才能得到令人满意的结果。

6. 步骤 6——实施解决

对模型及其解分析和检验后，还要对其实际意义给予解释并将其提交给管理决策者。如果管理决策者认为可以实施，则管理科学的分析结果得以实现，就能利用获得的正确模型对研究的实际问题给出预报或对类似实际问题进行分析、解释和预报，以供决策者参考；如果认为仍不能实施，则要再次对问题进行分析并适当修改模型。该步骤是运筹学分析结果回归管理环境的过程，也是运筹技术与管理科学的接口-出口过程。

运筹学的方法与技术能起到辅助决策的作用，但不能代替管理者的决策，管理者在分析各方面的因素和作出决策方面负有最终的责任。

1.4　本书的组织框架

本书由绪论、线性规划、图与网络、决策技术、排队理论和存储技术等 9 章组成，其框架如图 1.5 所示。第 1 章是全书的导引，其他各章相对独立，自成体系。

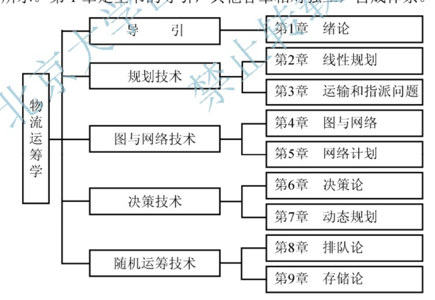


图 1.5　本书的组织框架

如今，各类组织及其赖以生存的客观世界都在以一种令人炫目的速度变得越来越复杂，无数的选择、无尽的时间压力和无止境边际利润的追求使得管理变得越来越难，新的企业管理手段为企业提供了巨量的管理数据，要把这些数据转化成对于组织未来的洞察力和行动计划是一件极具挑战性、决定性的工作。正是这些数据的存在以及计算能力的不断增强，为决策者提供了一个重要的机遇——运筹技术终于能够摆脱数据能力有限的束缚来为管理实践服务。

北京大学出版社版权所有　禁止转载

运筹思想已经渗透到经济、管理和生活中的各个层面，在纷繁复杂的现象中展现出独特的魅力。运筹方法在物流管理中的应用极具潜力，可以预见运筹学新的思想、观点和方法会不断地涌现。本书探讨的这些运筹方法在物流活动中都有现实的应用背景，是从事物流管理活动所要求掌握的知识。

本 章 小 结

本章首先从管理科学的发展历程切入，介绍了管理科学的特点与应用，引出了管理科学的重要分支物流管理；其次阐述了物流管理的主要内容和特点，论证了物流管理与运筹学的必然联系；最后就运筹学的发展过程，运筹学的主要分支，运筹学的数学模型作了概括性的回顾，并总结了本章的重点内容——运筹学的工作程序。



关键术语(中英文)

管理科学(Management Science)　　运筹学(Operations Research)
物流管理(Logistics Management)　　系统(System)



知识链接

管理科学学派

管理科学学派也称计量管理学派、数量学派。也有人把计量管理学派与运筹学看成是统一术语，这是因为该学派正式成立于 1939 年由美国曼彻斯特大学教授布莱克特领导的运筹学小组。埃尔伍德·斯潘赛·伯法是西方管理科学学派的代表人物之一。这个学派认为，解决复杂系统的管理决策问题，可以用电子计算机作为工具，寻求最佳计划方案，以达到企业的目标。管理科学其实就是管理中的一种数量分析方法，它主要用于解决能以数量表现的管理问题，其作用在于通过管理科学的方法，减少决策中的风险，提高决策的质量，保证投入的资源发挥最大的经济效益。

就管理科学的实质而言，它是泰罗的科学管理的继续与发展，因为他们都力图抛弃凭经验、凭主观判断来进行管理，而提倡采用科学的方法，探求最有效的工作方法或最优方案，以达到最高的工作效率，以最短的时间、最小的支出，得到最大的效果。不同的是，管理科学的研究，已经突破了操作方法、作业研究的范围，而向整个组织的所有活动方面扩展，要求进行整体性的管理。由于现代科学技术的发展，一系列的科学理论和方法被引进到管理领域。因此，管理科学可以说是现代的科学管理。

基于管理科学的特征，大多数管理学家认为管理科学只是一种有效的管理方法，而不是一种管理学派，它仅适用于解决特定的管理问题，目前完全采用管理科学的定量方法来解决复杂环境下的组织问题还面临着许多实际困难。管理科学学派重点研究的是操作方法和作业方面的管理问题，注意管理中应用的先进工具和科学方法，对管理中人的作用注意甚少，这是它的不足之处。

北京大学出版社版权所有
禁止转载



习题1

1.1 填空题

(1) 运筹学的主要分支包括__________、__________、__________、__________、__________、__________等。

(2) 现代管理是以__________、__________、__________为理论基础。

1.2 判断题(正确的打√，错误的打×)

(1) 用运筹学分析与解决问题的过程是一个科学决策过程。 (　　)

(2) 运筹学作为一门现代的新兴科学，起源于第二次世界大战的军事活动。 (　　)

(3) 运筹学中所使用的模型有实物模型、图表模型、数学模型。 (　　)

(4) 用运筹学解决问题的核心是建立数学模型并对模型求解。 (　　)

(5) 运筹学是近代形成的一门管理科学。 (　　)

1.3 物流管理的内涵与特点是什么？

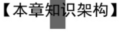
1.4 简要回顾运筹学的发展历史。

1.5 运用运筹学解决管理实际问题的步骤有哪些？

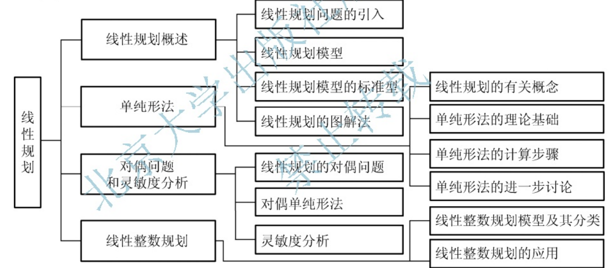
1.6 查阅文献，回答运筹学的数学模型有哪些优缺点？

北京大学出版社版权所有
禁止转载

第2章　线性规划



【本章知识架构】



北京大学出版社版权所有 禁止转载

【本章教学目标与要求】

- 掌握线性规划模型的三个要素，熟悉将一般线性规划问题转化为标准型，能够应用图解法分析两个变量线性规划解的情况。
- 理解单纯形法的思想，能应用单纯形表求解线性规划。
- 掌握线性规划对偶问题的建立方法，理解对偶问题的经济意义，掌握对偶单纯形算法，能对简单的线性规划进行灵敏度分析。
- 了解线性整数规划的分类，能够应用线性整数规划描述和分析实际问题。



运储物流配送中心的选址

合理地布局配送中心可以提高物流效率、降低物流成本。运储物流配送中心选址时除了要考察交通、土地、自然、行政等因素外，还要考虑以下因素。

(1) 客户的分布。运储物流的客户大部分是分布在人口密集的地方或大城市，配送中心为了提高服务水准及降低配送成本，其配送中心多建在城市边缘接近客户分布的地区。

(2) 供应商的分布。供应商的分布地区，因为物流的商品全部是由供应商所供应的，如果物流愈接近供应商，则其商品的安全库存可以控制在较低的水平。

经过以上综合分析，运储物流公司在广州、上海、北京、沈阳、合肥、成都、武汉等地设立了配送中心，以覆盖各个相邻区域的配送业务。

资料来源：http://shscyxs.mofcom.gov.cn.

线性规划分析是在满足企业内、外部的条件下，实现管理目标的一类问题，诸如国民经济的综合平衡、生产力的合理布局、最优计划与合理调度等问题。线性规划的广泛应用，除了它本身具有实用的特点之外，还由于线性规划模型的结构简单，比较容易被熟悉业务的经营管理人员所掌握。

北京大学出版社版权所有 禁止转载

2.1　线性规划概述

线性规划的研究内容可归纳为两个方面：一是资源的数量已定，如何合理利用、调配，使任务完成得最多，以便更有效地运用有限的资源，以更高水平达到目标；二是系统的任务已定，如何合理筹划，精细安排，用最少的资源(人力、物力和财力)去实现这项任务。

2.1.1　线性规划问题的引入

要想在生产、交通运输、商业贸易等领域提高经济效益，有两种途径：一是进行技术创新；二是提高管理水平，改进生产组织与计划，最合理地安排各类生产要素。

【例 2.1】(生产计划问题)某企业生产 1、2 和 3 三种型号的产品，每种产品需要经过三道工序，每件产品在每道工序中的工时定额、每道工序在每周可利用的有效工时和每件产品的利润见表 2-1。问每种产品各生产多少，可使这一周内生产的产品所获利润最大？

表 2-1　生产计划资源

定额/(工时/件)		产品型号 1	产品型号 2	产品型号 3	每周可利用的有效工时
工　序	*A*	1.2	1.0	1.1	5 400
	B	0.7	0.9	0.6	2 800
	C	0.9	0.8	1.0	3 600
利润/(元/件)		10	15	12	

解：该问题主要是把有限的工时资源合理地分配到三种产品的生产活动上去，以期望获得最多的利润。将工时资源安排多少到各工序是企业要作的决策，表示安排多少的变量就是决策变量。首先引进决策变量，设一周内企业各产品的生产件数为 $x_j\ (j=1,2,3)$；然后根据每件产品的工时定额以及各工序允许的有效工时的列出要求，产品 1 每生产一件需 A 工序 1.2 工时，现生产 x_1 件，故产品 1 耗费 A 工序的工时数为 $1.2x_1$；类似地，生产产品 2 和 3 所耗费 A 工序的工时数分别为 $1.0x_2$ 和 $1.1x_3$，这样三种产品对 A 工序的工时总需求量为 $1.2x_1+1.0x_2+1.1x_3$，它不应超过工序 A 在一周内所允许的工作时间 5 400 工时，于是，得到工序 A 加工产品的约束条件：$1.2x_1+1.0x_2+1.1x_3\leqslant 5\,400$。

同理，对工序 B 和 C 有以下约束条件：

$$0.7x_1+0.9x_2+0.6x_3\leqslant 2\,800$$
$$0.9x_1+0.8x_2+1.0x_3\leqslant 3\,600$$

再者，还有一些变量本身的约束，$x_j\,(j=1,2,3)$ 只能取非负值，故有下列非负约束条件：

$$x_1\geqslant 0,\quad x_2\geqslant 0,\quad x_3\geqslant 0$$

最后确定产品生产的效益，若用 z 表示工厂一周内生产三种产品所能获得的利润，则有

$$z=10x_1+15x_2+12x_3$$

根据问题的要求，旨在获得最大利润，也就是说，在资源约束的条件下，尽可能生产更多的产品，以获得最大的利润，实现工厂利润最大化的目标，因而目标可以表示为

$$\max z=10x_1+15x_2+12x_3$$

综上所述，得到生产计划问题的数学模型为

$$\max z=10x_1+15x_2+12x_3$$
$$\text{s.t.}\begin{cases}1.2x_1+1.0x_2+1.1x_3\leqslant 5\,400\\0.7x_1+0.9x_2+0.6x_3\leqslant 2\,800\\0.9x_1+0.8x_2+1.0x_3\leqslant 3\,600\\x_j\geqslant 0\qquad (j=1,2,3)\end{cases}$$

其中 s.t. 是英文 Subject to(受约束于)的缩写，也可以省略。

【例 2.2】(营养配餐问题)假定一个成年人每天需要从食物中获取 3 000cal 热量(1cal=4.1840J)，55g 蛋白质和 800mg 钙。如果市场上只有 4 种食品可选，它们每千克所含热量、营养构成，以及市场价格见表 2-2。问如何选择各种食品，才能满足成年人每天的营养需要，同时使购买食品的费用最小？

表 2-2　食品所含热量、营养构成以及市场价格

序　号	食品名称	热量/cal	蛋白质/g	钙/mg	价格/元
1	猪肉	1 000	50	400	10
2	鸡蛋	800	60	200	6
3	大米	900	20	300	3
4	白菜	200	10	500	2

解：该问题是如何选择各种食品的数量，以期在满足营养的前提下使购买食品的费用最小。首先引进决策变量，令 $x_j\ (j=1,2,3,4)$ 表示第 j 种食品每天的购买量；然后根据每

北京大学出版社版权所有
禁止转载

天从食物中获得营养的要求，猪肉中含有热量 1 000，现购买量为 x_1，故从猪肉中获得的热量为 $1\,000x_1$；类似地，从鸡蛋、大米和白菜中获得的热量分别为 $800x_2$、$900x_3$ 和 $200x_4$，这样从食品中获得的总热量为 $1000x_1+800x_2+900x_3+200x_4$，它应超过成人一天中所需的热量要求 3 000，于是，得到约束 $1\,000x_1+800x_2+900x_3+200x_4 \geqslant 3\,000$。

同理，蛋白质和钙要求的约束：

$$50x_1+60x_2+20x_3+10x_4 \geqslant 55$$

$$400x_1+200x_2+300x_3+500x_4 \geqslant 800$$

再者，还有一些变量本身的约束，$x_j\,(j=1,2,3)$ 只能取非负值，故有下列非负约束：

$$x_1 \geqslant 0,\ x_2 \geqslant 0,\ x_3 \geqslant 0,\ x_4 \geqslant 0$$

最后确定购买这食品的费用，若用 z 表示购买食品的总费用，则有

$$z=10x_1+6x_2+3x_3+2x_4$$

根据问题的要求，旨在总的费用最小，这是该问题的目标，可以表示为

$$\min z=10x_1+6x_2+3x_3+2x_4$$

综上所述，得配餐问题数学模型为

$$\min z=10x_1+6x_2+3x_3+2x_4$$

$$\text{s.t.}\begin{cases}1\,000x_1+800x_2+900x_3+200x_4 \geqslant 3\,000\\ 50x_1+60x_2+20x_3+10x_4 \geqslant 55\\ 400x_1+200x_2+300x_3+500x_4 \geqslant 800\\ x_j \geqslant 0 \qquad (j=1,2,3,4)\end{cases}$$

北京大学出版社版权所有 禁止转载

2.1.2 线性规划模型

2.1.1 节建立的数学模型称之为线性规划，简记为 LP。LP 模型具有以下 3 个要素。

(1) 决策变量。这些决策变量的一组定值代表所给问题的一个具体解决方案。一般来说，这些决策变量都是非负变量，如果在模型中变量的符号不受限制，即变量可以取正值、也可以取负值或零，那么将它称为自由变量，可以写入规划模型，也可以省略。

(2) 约束条件。这些约束条件都为线性等式或线性不等式。它们反映了所给问题对资源的客观限制及对所要完成的任务的各类要求，对决策变量的符号要求也属于约束条件。

(3) 目标函数。它为决策变量的线性函数。基于所给问题的不同，可要求目标函数实现最大值或最小值。

包含以上三个要素的数学模型称为线性规划的数学模型，其一般形式为

$$\max(\text{或}\min) z=c_1x_1+c_2x_2+\cdots+c_nx_n \tag{2-1}$$

$$\begin{cases}a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n \leqslant(=,\geqslant)b_1\\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n \leqslant(=,\geqslant)b_2\\ \qquad\vdots \\ a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n \leqslant(=,\geqslant)b_m \qquad (2\text{-}2)\\ x_1,x_2,\cdots,x_n \geqslant 0 \qquad (2\text{-}3)\end{cases}$$

矩阵形式为

$$\max(\text{或}\min)z = \boldsymbol{CX}$$
$$\begin{cases} \boldsymbol{AX} \leqslant (=, \geqslant)\boldsymbol{b} \\ \boldsymbol{X} \geqslant \boldsymbol{0} \end{cases} \tag{2-4}$$

向量形式为

$$\max(\text{或}\min)z = \boldsymbol{CX}$$
$$\begin{cases} \sum_{j=1}^{n} \boldsymbol{P}_j x_j \leqslant (=, \geqslant)\boldsymbol{b} \\ x_j \geqslant 0 \quad j = 1, 2, \cdots, n \end{cases} \tag{2-5}$$

其中：$\boldsymbol{C} = (c_1, c_2, \cdots, c_n)$ 是 n 维行向量，习惯上称为价值系数向量，简称为价值向量；$\boldsymbol{b} = (b_1, b_2, \cdots, b_m)^{\mathrm{T}}$ 是 m 维列向量，称为资源向量；$\boldsymbol{X} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 称为决策向量；$\boldsymbol{A}$ 为技术系数矩阵(也称为消耗系数矩阵)，简称为技术矩阵。

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = (\boldsymbol{P}_1 \ \boldsymbol{P}_2 \ \cdots \ \boldsymbol{P}_n)，\quad \boldsymbol{P}_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix}$$

以上三种线性规划的形式在后面的学习中都会用到，哪一种表示更清楚、简洁，就采用哪种形式。建立线性规划模型就是为了确定决策变量的取值，找出模型满足目标函数要求的最优解，为形成合理的计划方案服务。

北京大学出版社版权所有
禁止转载

2.1.3 线性规划模型的标准型

为便于讨论线性规划问题的解，一般把目标函数最大化、约束条件为等式、变量符号为非负的线性规划称为标准型，即

$$\max z = \boldsymbol{CX}$$
$$\begin{cases} \boldsymbol{AX} = \boldsymbol{b} \\ \boldsymbol{X} \geqslant \boldsymbol{0} \end{cases}$$

线性规划的标准型的特点：①目标函数是最大化类型；②约束条件均由等式组成；③决策变量均为非负。

对于各种非标准形式的线性规划都可以通过适当的变换转化为等价的标准型问题，具体包括以下做法。

1. 目标函数的调整

如果线性规划目标函数为最小，通过改变目标函数的符号，然后求最大。例如原问题 $\min z = 3x_1 + 5x_2$，只需要将所求目标函数的最小值变换为求另一个目标函数的最大值，就可化为 $\max z' = -3x_1 - 5x_2$。这样得到的新问题与原问题具有同样的可行域和最优解(若存在)，只是最优值(若存在)相差一个符号而已。

2. 约束条件的调整

(1) 如果线性规划具有“≤”不等式的约束，这时可引入一个松弛变量[①]变为等式约束。例如对约束 $4x_1+2x_2\leqslant 14$，加上松弛变量 $x_3\geqslant 0$，构成等式约束 $4x_1+2x_2+x_3=14$，同时规定目标函数中的价值系数 $c_3=0$，不改变原问题的最优解。

(2) 如果线性规划具有“≥”不等式的约束，这时可引入一个剩余变量[②]变为等式约束。例如对约束 $4x_1+2x_2\geqslant 14$，减去剩余变量 $x_3\geqslant 0$，构成等式约束 $4x_1+2x_2-x_3=14$，同时规定目标函数中的价值系数 $c_3=0$。

3. 决策变量的调整

(1) 如果决策变量为 $x_j<0$，引进新的非负变量 x_j'，令 $x_j=-x_j'$，代入约束条件和目标函数中消去 x_j，这样 $x_j<0$ 就化成 $x_j'\geqslant 0$。

(2) x_j 为自由变量，这时需引入两个新的非负变量 x_j'、x_j''，令 $x_j=x_j'-x_j''$，将其代入约束条件和目标函数中，消去 x_j，同时在约束条件中加入约束 x_j'、$x_j''\geqslant 0$。

将一个非标准线性规划调整为标准型的考虑顺序为：①决策变量；②约束条件；③目标函数。

【例 2.3】将下面线性规划问题化成标准型。

$$\min z=x_1+2x_2+3x_3$$
$$\begin{cases}-2x_1+x_2+x_3\leqslant 9\\ -3x_1+x_2+2x_3\geqslant 4\\ 4x_1-2x_2-3x_3=-6\\ x_1\leqslant 0,\quad x_2\geqslant 0,\quad x_3\text{无约束}\end{cases}$$

解：该线性规划共有三处不符合标准型要求，目标函数 z 求最小值；第一、二个约束条件为不等式；决策变量 $x_1\leqslant 0$；x_3 无约束。为此，通过以下步骤将该模型标准化。

(1) 决策变量 $x_1=-x_1'$，$x_3=x_3'-x_3''$。

(2) 第一个约束引入松弛变量 x_4，第二个约束引入剩余变量 x_5。

(3) 目标函数 min 变为 max。

由于在第三个约束中资源量为-6，也可以从实际意义考虑将第三个等式约束两边同乘以-1，按照上述步骤，该问题的标准形式为

$$\max z'=x_1'-2x_2-3x_3'+3x_3''+0x_4+0x_5$$
$$\begin{cases}2x_1'+x_2+x_3'-x_3''+x_4=9\\ 3x_1'+x_2+2x_3'-2x_3''-x_5=4\\ 4x_1'+2x_2+3x_3'-3x_3''=6\\ x_1',x_2,x_3',x_3'',x_4,x_5\geqslant 0\end{cases}$$

---

① 松弛变量表示在一个决策过程中资源消耗的剩余量。若为正，表示有剩余；若为零，表示没有剩余，其结果不影响收入，也不影响支出。因此，松弛变量本身是零价格的，表现在目标函数中，松弛变量的系数为零。

② 剩余变量表示在一个经济决策中超额满足最低需求的量。若为正，表示超额满足最低需求；若为负，表示没有满足最低需求，因此，剩余变量本身也是零价格的。

北京大学出版社版权所有 禁止转载

注意求解所得最优值的相反数才是原问题的最优值，从而各种各样的线性规划问题都可转化为标准型。

2.1.4 线性规划的图解法

对于只有两个变量的线性规划问题，可以直接使用图解法求解。图解法简单直观，对一般线性规划问题的求解富有启发作用。

图解法可分以下 3 步进行。

(1) 在平面上建立直角坐标系。

(2) 根据约束条件画出相应的半平面，由这些半平面共同确定的区域即为可行域。

(3) 画出目标函数的等值线，然后沿目标函数要求的方向平移至与可行域边界相切的点，该点就是最优解，相应的坐标即为最优解。

【例 2.4】求解线性规划问题。

$$\max z = 40x_1 + 50x_2$$
$$\begin{cases} x_1 + 2x_2 \leqslant 30 \\ 3x_1 + 2x_2 \leqslant 60 \\ 2x_2 \leqslant 24 \\ x_1, x_2 \geqslant 0 \end{cases}$$

解：应用图解法求解线性规划，分以下 3 步求解。

(1) 建立直角坐标系。取决策变量 x_1、x_2 为坐标向量建立直角坐标系，在直角坐标系中，图上任意一点的坐标代表了决策变量的一组取值。

(2) 确定可行域。可行域如图 2.1 所示。令 3 个约束条件为等式，得到 3 条直线，画出第一象限的三个不等式区域，它们的交集就是可行域，即区域 $OABCDO$，其内部及边界上的每一个点都是线性规划的解。

(3) 确定最优解。目标函数的等值线 $z = 40x_1 + 50x_2$ (z 取定某一个常值)的法线方向 $(40, 50)$ (梯度方向)是函数值增加最快的方向。沿着函数的梯度方向移动(图 2.1 箭头方向)，函数值会增大，当移动到顶点 $C(15, 15/2)$ 时，再继续移动就离开区域了。于是 C 点就是最优解，最优值为 $z = 40\times15 + 50\times15/2 = 975$。

【例 2.5】求解线性规划问题。

$$\max z = 40x_1 + 80x_2$$
$$\begin{cases} x_1 + 2x_2 \leqslant 30 \\ 3x_1 + 2x_2 \leqslant 60 \\ 2x_2 \leqslant 24 \\ x_1, x_2 \geqslant 0 \end{cases}$$

解：该题将例 2.4 中的目标函数变为 $z = 40x_1 + 80x_2$，可行域与例 2.4 的线性规划相同，用图解法求解的过程如图 2.2 所示。由于目标函数的 $z = 40x_1 + 80x_2$ 等值线与直线 $x_1 + 2x_2 = 30$ 平行，当目标函数的等值线与直线 $x_1 + 2x_2 = 30$ 重合时，目标函数 $z = 40x_1 + 80x_2$ 达到最大值 1 200，于是，线段 BC 上的每一个点均为该问题的最优解。特别地，线段 BC 的两个端点，即可行区域的两个顶点 $B(6,12)$、$C(15,15/2)$，均是该线性规划问题的最优解，此时，最优解不唯一。

北京大学出版社版权所有 禁止转载

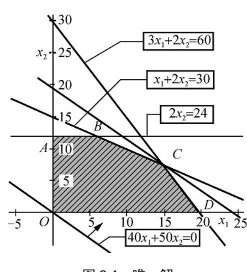


图 2.1 唯一解

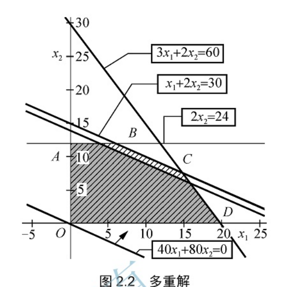


图 2.2 多重解

【例 2.6】求解线性规划问题。

$$\max z = 2x_1 + 4x_2$$

$$\begin{cases} 2x_1 + x_2 \geqslant 8 \\ -2x_1 + x_2 \leqslant 2 \\ x_1, x_2 \geqslant 0 \end{cases}$$

解：与例 2.5 求解方法类似，目标函数 $z = 2x_1 + 4x_2$ 沿着它的法线方向 $(2, 4)$ 移动，由于可行域无界(图 2.3)，因此，移动可以无限制下去，而目标函数值一直增加，所以该线性规划问题无有限最优解，有无界解，即该问题的目标函数无界。

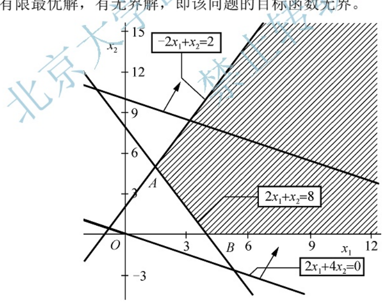


图 2.3 无界解

【例 2.7】求解线性规划问题。

$$\max z = 4x_1 + 2x_2$$

$$\begin{cases} -x_1 - 2x_2 \geqslant 2 \\ x_1, x_2 \geqslant 0 \end{cases}$$

解：约束条件互不相容，没有公共交点，可行域是一个空集，不存在满足条件的解，说明线性规划问题无解，没有最优解，如图 2.4 所示。

北京大学出版社版权所有 禁止转载

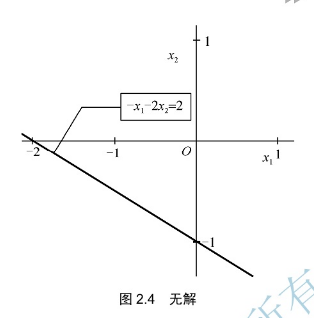


图 2.4　无解

综上所述，容易得到下面几个重要结论。

(1) 线性规划的可行域是若干个半平面的交集，它形成了一个多面凸集(可能是空集)。

(2) 对于给定的线性规划问题，如果它有最优解，最优解总可以在可行域的某个顶点上达到。这种情况下包含两种解的情况：唯一解和无穷多解。

(3) 如果可行域无界，线性规划问题的目标函数可能无界。

(4) 可行域为空集，无最优解。

北京大学出版社版权所有 禁止转载

2.2　单 纯 形 法

单纯形法是解决线性规划问题的基本方法，其原理出自线性代数的矩阵理论，是可按现代计算机标准程序求解线性规划模型的一般方法，分为代数形式的单纯形法和表格形式的单纯形法。前者提供基本算法所依据的逻辑规则，适用于在计算机上求解运算；后者将变量和数据列成表格，是本章采用的方法。

2.2.1　线性规划的有关概念

设线性规划的标准型为

$$\max z = \boldsymbol{CX}$$
$$\begin{cases} \boldsymbol{AX} = \boldsymbol{b} \\ \boldsymbol{X} \geqslant \boldsymbol{0} \end{cases}$$

其中：$\boldsymbol{A}$ 是 $m \times n$ 矩阵，$m \leqslant n$ 且秩 $r(\boldsymbol{A}) = m$，即 $\boldsymbol{A}$ 中至少有一个 $m \times m$ 满秩子矩阵。

为便于叙述线性规划求解过程，引入线性规划的有关概念。

解：满足约束条件(2-2)的决策向量 $\boldsymbol{X} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 称为线性规划的解。

可行解：满足条件(2-3)的解 $\boldsymbol{X} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 称为线性规划的可行解。

可行域：全体可行解的集合称为线性规划的可行域，一般记作 $D = \{\boldsymbol{X} \mid \boldsymbol{AX} = \boldsymbol{b}, \boldsymbol{X} \geqslant \boldsymbol{0}\}$。

最优解：使目标函数达到最大值(或最小值)的可行解 $\boldsymbol{X}^* = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 称为线性规划的最优解。

最优值：最优解的目标函数值称为线性规划的最优值。

基：若 $\boldsymbol{A}$ 中 $m\times m$ 子矩阵 $\boldsymbol{B}$ 满足 $r(\boldsymbol{B})=m$，则称 $\boldsymbol{B}$ 是线性规划问题的一个基。也就是说，矩阵 $\boldsymbol{B}$ 是由 $\boldsymbol{A}$ 的 m 个线性无关列向量所构成，不妨设为

$$\boldsymbol{B}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & & \vdots & \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix}=(\boldsymbol{P}_1 \quad \boldsymbol{P}_2 \quad \cdots \quad \boldsymbol{P}_m)$$，称 $\boldsymbol{P}_j(j=1,2,\cdots,m)$ 为基向量。

基变量、非基变量：与基向量 $\boldsymbol{P}_j(j=1,2,\cdots,m)$ 相对应的变量 $x_j(j=1,2,\cdots,m)$ 为基变量，其他决策变量称为非基变量。

基本解：记基变量为 $\boldsymbol{X}_B=(x_1,x_2,\cdots,x_m)^{\mathrm{T}}$，非基变量为 $\boldsymbol{X}_N=(x_{m+1},x_{m+2},\cdots,x_n)^{\mathrm{T}}$，称满足方程组 $\boldsymbol{B}\boldsymbol{X}_B=\boldsymbol{b}$ 且 $\boldsymbol{X}_N=0$ 的解为基本解。

基本可行解：若基本解 $\boldsymbol{X}_B=\boldsymbol{B}^{-1}\boldsymbol{b}\geqslant\boldsymbol{0}$，则称该解为基本可行解，$\boldsymbol{B}$ 为可行基。

线性规划解之间的关系如图 2.5 所示。容易验证基本可行解一定是可行解，后面将指出基本可行解是可行域中特殊的解。

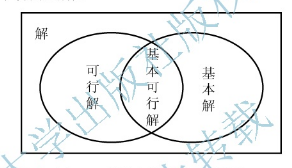


图 2.5 线性规划解之间的关系

北京大学出版社版权所有 禁止转载

【例 2.8】求出下面线性规划的所有基本解，并指出哪些是基本可行解。

$$\max z=2x_1+x_2$$
$$\begin{cases} 3x_1+5x_2\leqslant 15 \\ 6x_1+2x_2\leqslant 24 \\ x_1,\ x_2\geqslant 0 \end{cases}$$

解：将已知线性规划模型化为标准型

$$\max z=2x_1+x_2$$
$$\begin{cases} 3x_1+5x_2+x_3=15 \\ 6x_1+2x_2+x_4=24 \\ x_1,\ x_2,\ x_3,\ x_4\geqslant 0 \end{cases}$$

其中：系数矩阵 $\boldsymbol{A}=\begin{bmatrix} 3 & 5 & 1 & 0 \\ 6 & 2 & 0 & 1 \end{bmatrix}=[\boldsymbol{P}_1 \quad \boldsymbol{P}_2 \quad \boldsymbol{P}_3 \quad \boldsymbol{P}_4]$，$r(\boldsymbol{A})=2$。

选取矩阵 $\boldsymbol{B}_1=\begin{bmatrix} 3 & 5 \\ 6 & 2 \end{bmatrix}$，由于 $\boldsymbol{B}_1$ 的行列式值不为零，所以 $r(\boldsymbol{B}_1)=2$，即 $\boldsymbol{B}_1$ 是已知线性规划的一个基，基变量是 (x_1,x_2)。令非基变量 (x_3,x_4) 为零，此时线性规划的约束条件变为

$\begin{cases}3x_1+5x_2=15\\6x_1+2x_2=24\end{cases}$，求解得 $\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}3&5\\6&2\end{bmatrix}^{-1}\begin{bmatrix}15\\24\end{bmatrix}=\begin{bmatrix}15/4\\3/4\end{bmatrix}$，相应的线性规划基本解为 $\boldsymbol{X}_1=(15/4,3/4,0,0)$。由于 $\boldsymbol{X}_1\geqslant\boldsymbol{0}$，$\boldsymbol{X}_1$ 也是基本可行解。

其他基本解见表 2-3。

表 2-3　基本解与基本可行解

基	基向量	基变量	非基变量	基本解	基本可行解
$\boldsymbol{B}_1=\begin{bmatrix}3&5\\6&2\end{bmatrix}$	$(\boldsymbol{P}_1,\boldsymbol{P}_2)$	(x_1,x_2)	(x_3,x_4)	$\boldsymbol{X}_1=(15/4,3/4,0,0)$	是
$\boldsymbol{B}_2=\begin{bmatrix}3&1\\6&0\end{bmatrix}$	$(\boldsymbol{P}_1,\boldsymbol{P}_3)$	(x_1,x_3)	(x_2,x_4)	$\boldsymbol{X}_2=(4,0,3,0)$	是
$\boldsymbol{B}_3=\begin{bmatrix}3&0\\6&1\end{bmatrix}$	$(\boldsymbol{P}_1,\boldsymbol{P}_4)$	(x_1,x_4)	(x_2,x_3)	$\boldsymbol{X}_3=(5,0,0,-6)$	不是
$\boldsymbol{B}_4=\begin{bmatrix}5&1\\2&0\end{bmatrix}$	$(\boldsymbol{P}_2,\boldsymbol{P}_3)$	(x_2,x_3)	(x_1,x_4)	$\boldsymbol{X}_4=(0,12,-45,0)$	不是
$\boldsymbol{B}_5=\begin{bmatrix}5&0\\2&1\end{bmatrix}$	$(\boldsymbol{P}_2,\boldsymbol{P}_4)$	(x_2,x_4)	(x_1,x_3)	$\boldsymbol{X}_5=(0,3,0,18)$	是
$\boldsymbol{B}_6=\begin{bmatrix}1&0\\0&1\end{bmatrix}$	$(\boldsymbol{P}_3,\boldsymbol{P}_4)$	(x_3,x_4)	(x_1,x_2)	$\boldsymbol{X}_6=(0,0,15,24)$	是

一般来说，如果线性规划具有 n 个变量 m 个约束，则基本解的数量小于或等于 C_n^m，当然基本可行解的数量更不会超过 C_n^m 个。

2.2.2　单纯形法的理论基础

简要概括单纯形法基本概念和基本定理。

1. 基本概念

凸集：设 D 是 n 维欧氏空间的一个点集，若任意两点 $x^{(1)}\in D,x^{(2)}\in D$ 的连线段上的一切点 $\alpha x^{(1)}+(1-\alpha)x^{(2)}\in D\ (0\leqslant\alpha\leqslant1)$，则称 D 为凸集。

凸组合：设 $x^{(1)}$，$x^{(2)},\cdots$，$x^{(k)}$ 是 n 维欧氏空间中的 k 个点，若存在 $u_1,u_2,\cdots,u_k$，且 $0\leqslant u_i\leqslant1\ (i=1,2,\cdots,k)$, $\sum_{i=1}^{k}u_i=1$ 使 $x=u_1x^{(1)}+u_2x^{(2)}+\cdots+u_kx^{(k)}$，则称 x 为 $x^{(1)}$，$x^{(2)},\cdots$，$x^{(k)}$ 的凸组合。

顶点：设 D 是凸集，若 x 不能用不同的两点 $x^{(1)}\in D,x^{(2)}\in D$ 的凸组合表示为 $x=\alpha x^{(1)}+(1-\alpha)x^{(2)}\in k\ (0<\alpha<1)$，则称 x 为 D 的一个顶点(或极点)。

2. 基本定理

定理 2.1 若线性规划问题存在可行域，则可行域是凸集。

定理 2.2 线性规划可行域 D 中的点 $\boldsymbol{X}$ 为顶点的充要条件是 $\boldsymbol{X}$ 为基本可行解。

定理 2.3 若线性规划有最优解，则最优解一定可以在可行域的顶点上得到。

北京大学出版社版权所有 禁止转载

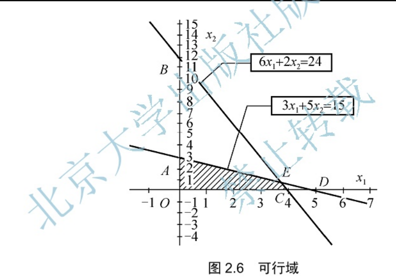
定理 2.1 描述了可行域的特征，定理 2.2 描述了可行域的顶点与基本可行解的对应关系，定理 2.3 描述了最优解在可行域中的位置。若最优解唯一，则最优解只能在某一顶点上达到(图 2.1)；若最优解不唯一，则最优解是某些顶点的凸组合，不可能是可行域的内点(图 2.2)。若可行域有界，则一定有最优解；若可行域为无界，则可能有最优解，也可能无最优解，若有也必定在某个顶点达到。

【例 2.9】 指出例 2.8 中的线性规划基本解与其可行域顶点的对应关系。

解：参见表 2-4 的说明及图 2.6 所示。

表 2-4　基本解与其对应的顶点

基 变 量	基 本 解	基本可行解	顶 点
(x_1, x_2)	$\boldsymbol{X}_1 = (15/4, 3/4, 0, 0)$	是	E
(x_1, x_3)	$\boldsymbol{X}_2 = (4, 0, 3, 0)$	是	C
(x_1, x_4)	$\boldsymbol{X}_3 = (5, 0, 0, -6)$	不是	D
(x_2, x_3)	$\boldsymbol{X}_4 = (0, 12, -45, 0)$	不是	B
(x_2, x_4)	$\boldsymbol{X}_5 = (0, 3, 0, 18)$	是	A
(x_3, x_4)	$\boldsymbol{X}_6 = (0, 0, 15, 24)$	是	O



北京大学出版社版权所有 禁止转载

图 2.6　可行域

线性规划问题的所有解构成的集合是凸集，凸集的每一个顶点对应一个基本可行解，由于基本可行解的个数是有限的(它不大于 C_n^m 个)，所以顶点的个数也是有限的。若线性规划问题有最优解，必在可行域某顶点上得到，可采用“枚举法”找出所有基本可行解，然后一一进行比较，最终就可能找到最优解。但当 n, m 较大时，计算工作烦琐庞大，这种方法还是行不通的，所以还要继续讨论如何有效地找到最优解的方法，这就是下面介绍的单纯形法。

2.2.3　单纯形法的计算步骤

单纯形法是从可行域的一个基本可行解出发(顶点)，判断该解是否为最优解，如果不是就转移到另一个较好的基本可行解，如果目标函数达到最优，则已得到最优解；否则，继续转移到其他较好的基本可行解。由于基本可行解(顶点)的数目是有限的，所以经过有限次数的迭代转换后就能求出最优解，如图 2.7 所示。

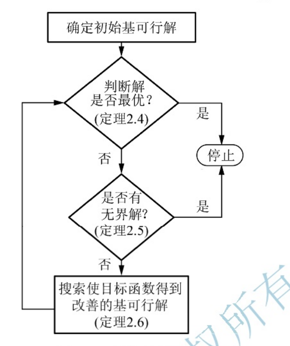


图 2.7　单纯形求解步骤

北京大学出版社版权所有
禁止转载

1. 初始基本可行解的确定

(1) 若给定问题标准化后(且 $\boldsymbol{b}\geqslant 0$)，系数矩阵 $\boldsymbol{A}$ 中存在 m 个线性无关的单位列向量，则以这 m 个单位列向量构成的单位矩阵作为初始基 $\boldsymbol{B}$，则 $\boldsymbol{X}_B=\boldsymbol{B}^{-1}\boldsymbol{b}=\boldsymbol{b}\geqslant 0$，其他 $x_j=0$ 就是初始基本可行解。

【例 2.10】求下面问题的初始基本可行解。

$$\max z=6x_1+4x_2$$

$$\begin{cases}2x_1+3x_2\leqslant 100\\ 4x_1+2x_2\leqslant 120\\ x_1,\quad x_2\geqslant 0\end{cases}$$

解：将已知线性规划标准化，添加松弛变量 x_3, x_4，得

$$\max z=6x_1+4x_2$$

$$\begin{cases}2x_1+3x_2+x_3=100\\ 4x_1+2x_2+x_4=120\\ x_1,x_2,x_3,x_4\geqslant 0\end{cases}$$

由于 $\boldsymbol{A}=\begin{pmatrix}2&3&1&0\\4&2&0&1\end{pmatrix}=(\boldsymbol{P}_1\quad \boldsymbol{P}_2\quad \boldsymbol{P}_3\quad \boldsymbol{P}_4)$ 有两个单位列向量，可取 $\boldsymbol{B}=\begin{pmatrix}1&0\\0&1\end{pmatrix}$ 为初始基，则

$$\boldsymbol{X}_B=\begin{pmatrix}x_3\\x_4\end{pmatrix}=\boldsymbol{B}^{-1}\boldsymbol{b}=\begin{pmatrix}100\\120\end{pmatrix},\quad \begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$$

就是初始基本可行解。

(2) 若给定问题标准化后($\boldsymbol{b}\geqslant 0$)系数矩阵中不存在 m 个线性无关的单位列向量，则在某些约束左端加一个非负的人工变量构建一些单位列向量，人工创造一个单位矩阵，参看后面的大 M 法。

2. 线性规划的求解定理

线性规划的标准型为

$$\max z = \boldsymbol{CX}$$
$$\text{s.t.} \begin{cases} \boldsymbol{AX} = \boldsymbol{b} \\ \boldsymbol{X} \geqslant \boldsymbol{0} \end{cases} \tag{2-6}$$

其中：$\boldsymbol{A}$ 是 $m \times n$ 矩阵且秩 $r(\boldsymbol{A}) = m$。不妨设 $\boldsymbol{B} = (\boldsymbol{P}_1 \quad \boldsymbol{P}_2 \quad \cdots \quad \boldsymbol{P}_m)$ 是线性规划的一个基，将有关矩阵和向量分块，记 $\boldsymbol{A} = (\boldsymbol{B}, \boldsymbol{N})$，$\boldsymbol{C} = (\boldsymbol{C}_B, \boldsymbol{C}_N)$，$\boldsymbol{X} = (\boldsymbol{X}_B, \boldsymbol{X}_N)^{\mathrm{T}}$，为求线性规划的基本最优解，先要求线性规划的基本可行解，这样就需将约束中的基变量用非基变量表示出来。

用 $\boldsymbol{B}^{-1}$ 左乘约束方程 $\boldsymbol{AX} = \boldsymbol{b}$ 的两端，得

$$\boldsymbol{B}^{-1}\boldsymbol{AX} = \boldsymbol{B}^{-1}(\boldsymbol{B}, \boldsymbol{N})\begin{pmatrix} \boldsymbol{X}_B \\ \boldsymbol{X}_N \end{pmatrix} = \boldsymbol{B}^{-1}\boldsymbol{BX}_B + \boldsymbol{B}^{-1}\boldsymbol{NX}_N = \boldsymbol{B}^{-1}\boldsymbol{b}$$

即
$$\boldsymbol{B}^{-1}\boldsymbol{BX}_B + \boldsymbol{B}^{-1}\boldsymbol{NX}_N = \boldsymbol{EX}_B + \boldsymbol{B}^{-1}\boldsymbol{NX}_N = \boldsymbol{B}^{-1}\boldsymbol{b} \tag{2-7}$$

其中 $\boldsymbol{E}$ 是单位矩阵。整理，得

$$\boldsymbol{X}_B = \boldsymbol{B}^{-1}\boldsymbol{b} - \boldsymbol{B}^{-1}\boldsymbol{NX}_N$$

将其代入目标函数中，得

$$z = \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b} - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{NX}_N + \boldsymbol{C}_N\boldsymbol{X}_N = \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b} + (\boldsymbol{C}_N - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{N})\boldsymbol{X}_N$$

即
$$z = \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b} + \sum_{j=m+1}^{n}(c_j - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{P}_j)x_j \tag{2-8}$$

非基变量 x_j 前面的系数 $c_j - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{P}_j$ 称为变量 x_j 的检验数，表示该变量增加或减少一个单位所引起目标函数值的变化，它可以用来判断将 x_j 变成基变量后能否改进目标函数值，以后记为 $\sigma_j = c_j - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{P}_j$。

以上过程也可以通过线性方程组消元实现。

考虑线性方程组 $\begin{cases} \boldsymbol{AX} = \boldsymbol{b} \\ z = \boldsymbol{CX} \end{cases}$，其变量为 $\begin{bmatrix} \boldsymbol{X} \\ z \end{bmatrix}$，为便于求解，整理得方程组 $\begin{cases} 0 \cdot z + \boldsymbol{AX} = \boldsymbol{b} \\ -z + \boldsymbol{CX} = \boldsymbol{0} \end{cases}$，其增广矩阵见表 2-5(a)，应用高斯消元法，求解线性方程组的解。

表 2-5(a)　增广矩阵

常数项	z	$\boldsymbol{X}_B$	$\boldsymbol{X}_N$	
b	0	$\boldsymbol{B}$	$\boldsymbol{N}$	(2-9)
0	−1	$\boldsymbol{C}_B$	$\boldsymbol{C}_N$	(2-10)

用 $\boldsymbol{B}^{-1}$ 左乘方程组(2-9)的两端，将 $\boldsymbol{X}_B$ 的系数化为单位矩阵，得表 2−5(b)。

表 2-5(b)　系数单位化

常数项	z	$\boldsymbol{X}_B$	$\boldsymbol{X}_N$	
$\boldsymbol{B}^{-1}\boldsymbol{b}$	0	$\boldsymbol{B}^{-1}\boldsymbol{B} = \boldsymbol{E}$	$\boldsymbol{B}^{-1}\boldsymbol{N}$	(2-11)
0	−1	$\boldsymbol{C}_B$	$\boldsymbol{C}_N$	

北京大学出版社版权所有 禁止转载

将方程组(2-11)左乘$-\boldsymbol{C}_B$加到方程(2-10)两边，得表 2-5(c)。

表 2-5(c)　化简

常数项	z	$\boldsymbol{X}_B$	$\boldsymbol{X}_N$	
$\boldsymbol{B}^{-1}\boldsymbol{b}$	0	$\boldsymbol{B}^{-1}\boldsymbol{B}=\boldsymbol{E}$	$\boldsymbol{B}^{-1}\boldsymbol{N}$	
$-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b}$	−1	$\boxed{\boldsymbol{C}_B-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{B}}=0$	$\boxed{\boldsymbol{C}_N-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{N}}$	(2-12)

注意式(2-12)中基向量$\boldsymbol{X}_B$、非基向量$\boldsymbol{X}_N$的系数，它们形式相似，当前者$\boldsymbol{C}_B-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{B}=0$时，后者即为检验数$\boldsymbol{C}_N-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{N}$的矩阵形式。也就是说，用消元法将目标函数中基变量的系数化为零的同时，就会得到非基变量的检验数。事实上，$\boldsymbol{C}_B-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{B}$也可看成基向量的检验数，应用消元法后，当基向量的检验数变为零时，非基变量的系数$\boldsymbol{C}_N-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{N}$就是其检验数。

从以上论述可知，计算检验数的方法有两种：一是通过公式$\sigma_j=c_j-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{P}_j$计算得出；二是应用消元法，将目标函数基向量的系数都化为零时，非基向量的系数就是检验数。

线性规划单纯形法求解有下述定理。

定理 2.4 (最优解判定定理)对某基本可行解$\boldsymbol{X}_B=\boldsymbol{B}^{-1}\boldsymbol{b}$其他$x_N=0$，若所有的$\sigma_j=\boldsymbol{c}_j-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{P}_j\leqslant 0$，则该解为最优解。

注意式(2-8)，要想让非基变量为零而不影响最优值，只有其前面的系数都非正，即检验数为非正，这时非基变量取值必为零；同样可以得出，当目标函数为最小要求时，检验数要求为非负。

以上求解的过程也能得出结论：如果$\boldsymbol{X}$是线性规划(2-6)的基本最优解，那么$-z=-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b}$，$z=\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b}$，即线性规划(2-6)的最优解为($\boldsymbol{B}^{-1}\boldsymbol{b}$,0)。

定理 2.5 (无界解判定定理)若对某可行基$\boldsymbol{B}$，存在$\sigma_k>0$且$\boldsymbol{B}^{-1}\boldsymbol{P}_k\leqslant 0$，则该线性规划问题有无界解。

定理 2.6 (换基迭代定理)

(1) 给定某可行基$\boldsymbol{B}$，对矩阵$\boldsymbol{A}$的行列来说，通过下面步骤：①确定进基变量：$\sigma_k=\max\{\sigma_j>0\}$，选择$k$列对应的非基变量$x_k$为进基变量；②按最小比值原则确定出基变量，$\theta=\min\left[\dfrac{(\boldsymbol{B}^{-1}\boldsymbol{b})_i}{(\boldsymbol{B}^{-1}\boldsymbol{P}_k)_i}\,|\,(\boldsymbol{B}^{-1}\boldsymbol{P}_k)_i>0\right]=\dfrac{(\boldsymbol{B}^{-1}\boldsymbol{b})_l}{(\boldsymbol{B}^{-1}\boldsymbol{P}_k)_l}$(最小值不唯一时任选一个行标作为$l$)，其中$(\boldsymbol{B}^{-1}\boldsymbol{b})_i$表示向量$\boldsymbol{B}^{-1}\boldsymbol{b}$的第$i$个元素，选择$l$行对应的基变量为出基变量，那么$\boldsymbol{P}_k$替换$\boldsymbol{B}$中出基变量所在的列后就能得到一个新的可行基$\boldsymbol{B}'$。

(2) 对线性方程组$\boldsymbol{AX}=\boldsymbol{b}$的增广矩阵实施初等行变换：①第$l$行元素同除以主元$a_{lk}$，使主元变为 1；②使主元所在列的其他元素变为 0，这样就得到可行基$\boldsymbol{B}'$对应的基本可行解，并且所对应的目标函数值增加$\theta\sigma_k$。

3. 单纯形表

为便于表达单纯形法的计算过程，上述计算过程可以在单纯形表上完成。每个可行基$\boldsymbol{B}$都可构造出一个单纯形表，一般将其化为单位矩阵后再列出单纯形表，见表 2-6。

北京大学出版社版权所有
禁止转载

表 2-6　单纯形表

C_B	X_B	$B^{-1}b$	c_1	c_2	…	c_m	c_{m+1}	…	c_n	θ
			x_1	x_2	…	x_m	x_{m+1}	…	x_n	
c_1	x_1	b_1	1	0	…	0	a_{1m+1}	…	a_{1n}	
c_2	x_2	b_2	0	1	…	0	a_{2m+1}	…	a_{1n}	
…	…	…	…	…	…	…	…	…	…	…
c_m	x_m	b_m	0	0	…	1	a_{mm+1}	…	a_{mn}	
检验数 σ			0	0	…	0	σ_{m+1}	…	σ_n	

表 2-6 中 X_B 一列记录基变量；C_B 一列跟踪基变量对应的价值系数；$B^{-1}b$ 为基本可行解；其余各列为 $B^{-1}P_j(j=1,2,\cdots,n)$，最后一行是每个变量对应的检验数。表 2-6 是表 2-5(a)的展开形式，大方框中元素就是表 2-3 中的增广矩阵，这样求解线性规划就可以在单纯形表中实现，单纯形表包含了线性规划求解的全部信息。

对于以 B 为可行基的单纯形表具有以下特点。

(1) 基变量的值 $X_B = B^{-1}b$ 。

(2) 表中各列为 $B^{-1}P_j$ $(j=1,2,\cdots,n)$ 。

(3) 检验数 $\sigma_j = c_j - C_B B^{-1} P_j$ 。

(4) 表中基变量的列是单位列向量，且基变量对应的检验数等于零。

(5) 由于可行基 B 化为单位矩阵，也可以认为初始基是单位矩阵，在后面的迭代中，初始基始终对应各表基的逆矩阵。

【例 2.11】用单纯形法求例 2.8 的最优解。

$$\max z = 2x_1 + x_2$$
$$\begin{cases} 3x_1 + 5x_2 + x_3 = 15 \\ 6x_1 + 2x_2 + x_4 = 24 \\ x_1,\ x_2,\ x_3,\ x_4,\ \geqslant 0 \end{cases}$$

解： 易知线性规划的标准型中有两个线性无关的列向量，它们构成的单位矩阵可以作为初始可行基，x_3, x_4 是基变量，$X_B = \begin{bmatrix} x_3 \\ x_4 \end{bmatrix} = B^{-1}b = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 15 \\ 24 \end{bmatrix} = \begin{bmatrix} 15 \\ 24 \end{bmatrix}$，得到初始可行解为 $X^{(0)} = (0, 0, 15, 24)$，目标函数值 $z^{(0)} = 0$，对应图 2.6 中的原点 O。

计算检验数

$$\sigma_1 = c_1 - C_B B^{-1} P_1 = 2 - \begin{bmatrix} 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 6 \end{bmatrix} = 2,$$

$$\sigma_2 = c_2 - C_B B^{-1} P_2 = 1 - \begin{bmatrix} 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ 2 \end{bmatrix} = 1,$$

北京大学出版社版权所有 禁止转载

填入单纯形表 2-7(a)。

表 2-7(a)　单纯形表

$\boldsymbol{C}_B$	$\boldsymbol{X}_B$	$\boldsymbol{B}^{-1}\boldsymbol{b}$	2	1	0	0	θ
			x_1	x_2	x_3	x_4	
0	x_3	15	3	5	1	0	$15/3=5$
0	x_4	24	[6]	2	0	1	$24/6=4$
检验数 σ			2	1	0	0	

由于非基变量的检验数大于零，根据定理 2.4，该基本可行解不是最优解；又由于 $\boldsymbol{B}^{-1}\boldsymbol{P}_1=(3,6)^{\mathrm{T}}$ 和 $\boldsymbol{B}^{-1}\boldsymbol{P}_2=(5,2)^{\mathrm{T}}$ 都大于零，根据定理 2.5，该线性规划没有无界解；这样就需搜索目标函数得到改善的可行解，根据定理 2.6，由于 $\sigma_1=\max\{\sigma_j>0\}$ (确定矩阵 $\boldsymbol{A}$ 中的列)，选择第一列对应的变量 x_1 为进基变量；由于 $\theta=\min\left[\frac{15}{3},\frac{24}{6}\right]=\frac{24}{6}=4$ (确定矩阵 $\boldsymbol{A}$ 中的行)，$l=2$，选择第二行对应的变量 x_4 为出基变量，用 $\boldsymbol{P}_1=[3\quad 6]^{\mathrm{T}}$ 替换基中的 $\boldsymbol{P}_4=[0\quad 1]^{\mathrm{T}}$，这样就得到新的可行基，其基变量为 x_3,x_1，非基变量为 x_2,x_4。

以矩阵 $\boldsymbol{A}$ 第一列和第二行的交叉元素 $a_{21}=6$ 为主元，在单纯形表以“[ ]”提示，应用消元法简化线性方程组的增广矩阵，得

$$\boldsymbol{X}_B=\begin{bmatrix}x_3\\x_1\end{bmatrix}=\boldsymbol{B}^{-1}\boldsymbol{b}=\begin{bmatrix}1&3\\0&6\end{bmatrix}^{-1}\begin{bmatrix}15\\24\end{bmatrix}=\begin{bmatrix}3\\4\end{bmatrix}$$

得到新的基本可行解为 $\boldsymbol{X}^{(1)}=(4,0,3,0)$，目标函数值 $z^{(1)}=z^{(0)}+\theta\sigma_1=2\times4=8$，对应图 2.6 中的点 C。

计算检验数

$$\sigma_2=c_2-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{P}_2=1-[0\quad 2]\begin{bmatrix}1&3\\0&6\end{bmatrix}^{-1}\begin{bmatrix}5\\2\end{bmatrix}=\frac{1}{3},$$

$$\sigma_4=c_4-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{P}_4=0-[0\quad 2]\begin{bmatrix}1&3\\0&6\end{bmatrix}^{-1}\begin{bmatrix}0\\1\end{bmatrix}=-\frac{1}{3},$$

填入单纯形表 2-7(b)。

表 2-7(b)　单纯形表

$\boldsymbol{C}_B$	$\boldsymbol{X}_B$	$\boldsymbol{B}^{-1}\boldsymbol{b}$	2	1	0	0	θ
			x_1	x_2	x_3	x_4	
0	x_3	3	0	4	1	−1/2	$3/4=0.75$
2	x_1	4	1	1/3	0	1/6	$4\Big/\frac{1}{3}=12$
检验数 σ			0	1/3	0	−1/3	

同理，得到新的基本可行解 $\boldsymbol{X}^{(2)}=(15/4,3/4,0,0)$，目标函数值 $z^{(2)}=33/4$，对应图 2.6 中的点 E。计算检验数，填入单纯形表 2-7(c)。

北京大学出版社版权所有 禁止转载

表 2-7(c) 单纯形表

C_B	X_B	$B^{-1}b$	2	1	0	0	θ
			x_1	x_2	x_3	x_4	
1	x_2	3/4	0	1	1/4	−1/8	
2	x_1	15/4	1	0	−1/12	5/24	
检验数 σ			0	0	−1/12	−7/24	

从表 2-8(c)可知此时所有的检验数都非正，根据定理 2.4 得到线性规划的最优解 $\boldsymbol{X}^{(2)}$ 和最优值 $z^{(2)}$，也就是说，线性规划在 E 点达到最优。

【例 2.12】用单纯形法求解线性规划。

$$\max z = 3x_1 + 5x_2 + x_4$$

$$\begin{cases} x_1 + x_2 + x_3 \leqslant 35 \\ 5x_1 + x_2 - x_4 \leqslant 12 \\ -x_3 + x_4 \leqslant 5 \\ x_1,\ x_2,\ x_3,\ x_4 \geqslant 0 \end{cases}$$

解：将已知线性规划标准化，添加松弛变量 x_5, x_6, x_7，得

$$\max z = 3x_1 + 5x_2 + 0x_3 + x_4 + 0x_5 + 0x_6 + 0x_7$$

$$\begin{cases} x_1 + x_2 + x_3 + x_5 = 35 \\ 5x_1 + x_2 - x_4 + x_6 = 12 \\ -x_3 + x_4 + x_7 = 5 \\ x_1,\ x_2,\ x_3,\ x_4,\ x_5,\ x_6,\ x_7 \geqslant 0 \end{cases}$$

由于技术矩阵中有一个单位矩阵可作为初始可行基，可以确定出线性规划的初始可行解，建立单纯形表，解题过程见表 2-8(a)。

表 2-8(a) 单纯形表

C_B	X_B	$B^{-1}b$	3	5	0	1	0	0	0	θ
			x_1	x_2	x_3	x_4	x_5	x_6	x_7	
0	x_5	35	1	1	1	0	1	0	0	35
0	x_6	12	5	[1]	0	−1	0	1	0	12
0	x_7	5	0	0	−1	1	0	0	1	—
检验数 σ			3	5	0	1	0	0	0	

由于有些非基变量的检验数大于零，根据定理 2.4，该基本可行解不是最优解，从单纯形表中也可以看出，该线性规划没有无界解，所以搜索能使目标函数得到改善的可行解，根据定理 2.6，选择进基变量为 x_2：$\sigma_2 = \max\{\sigma_j > 0\}$。选择出基变量为 x_6：$\theta = \min\left[\frac{35}{1},\ \frac{12}{1}\right] = \frac{12}{1} = 12$ 和 $l = 2$，用 $\boldsymbol{P}_2 = [1\ \ 1\ \ 0]^{\mathrm{T}}$ 替换基中的 $\boldsymbol{P}_6 = [0\ \ 1\ \ 0]^{\mathrm{T}}$，得到新的可行基。以 $a_{22} = 1$ 为主元，应用消元法简化线性方程组的增广矩阵，同时用消元法计算其检验数(参看线性规划的求解定理)，得出单纯形表 2-8(b)。

北京大学出版社版权所有 禁止转载

表 2-8(b)　单纯形表

C_B	X_B	$B^{-1}b$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	θ
0	x_5	23	−4	0	1	1	1	−1	0	23
5	x_2	12	5	1	0	−1	0	1	0	—
0	x_7	5	0	0	−1	[1]	0	0	1	5
检验数 σ			−22	0	0	6	0	−5	0	
0	x_5	18	−4	0	[2]	0	1	−1	−1	9
5	x_2	17	5	1	−1	0	0	1	1	—
1	x_4	5	0	0	−1	1	0	0	1	—
检验数 σ			−22	0	6	0	0	−5	−6	
0	x_3	9	−2	0	1	0	1/2	−1/2	−1/2	
5	x_2	26	3	1	0	0	1/2	1/2	1/2	
1	x_4	14	−2	0	0	1	1/2	−1/2	1/2	
检验数 σ			−10	0	0	0	−3	−2	−3	

最优解 $\boldsymbol{X}^* = (0, 26, 9, 14, 0, ,0, 0)^{\mathrm{T}}$，最优值 $z^* = 144$。

利用单纯形表可清楚地计算出基变量的值和非基变量的检验数，不仅如此，使用单纯形表更易于进行换基迭代运算。

【例 2.13】用单纯形法求解线性规划。

$$\max z = 2x_1 + 4x_2$$

$$\begin{cases} -x_1 + 2x_2 \leqslant 4 \\ x_1 + 2x_2 \leqslant 10 \\ x_1 - x_2 \leqslant 2 \\ x_1, x_2 \geqslant 0 \end{cases}$$

解：将线性规划标准化，添加松弛变量 x_3, x_4, x_5，得

$$\max z = 2x_1 + 4x_2 + 0x_3 + 0x_4 + 0x_5$$

$$\begin{cases} -x_1 + 2x_2 + x_3 = 4 \\ x_1 + 2x_2 + x_4 = 10 \\ x_1 - x_2 + x_5 = 2 \\ x_1, x_2, x_3, x_4, x_5 \geqslant 0 \end{cases}$$

由于技术矩阵中有一个单位矩阵可作为初始可行基，可以确定出线性规划的初始可行解，建立单纯形表，解题过程见表 2-9。

表 2-9　单纯形表

C_B	X_B	$B^{-1}b$	2	4	0	0	0	θ
			x_1	x_2	x_3	x_4	x_5	
0	x_3	4	−1	[2]	1	0	0	2
0	x_4	10	1	2	0	1	0	5
0	x_5	2	1	−1	0	0	1	—
检验数 σ			2	4	0	0	0	

北京大学出版社版权所有
禁止转载

续表

C_B	X_B	$B^{-1}b$	2	4	0	0	0	θ
			x_1	x_2	x_3	x_4	x_5	
4	x_2	2	−1/2	1	1/2	0	0	—
0	x_4	6	[2]	0	−1	1	0	3
0	x_5	4	1/2	0	1/2	0	1	8
检验数 σ			4	0	−2	0	0	
4	x_2	7/2	0	1	1/4	1/4	0	
2	x_1	3	1	0	−1/2	1/2	0	
0	x_5	5/2	0	0	3/4	−1/4	1	
检验数 σ			0	0	0	−2	0	

最优解 $\boldsymbol{X}_1^* = (3, 7/2, 0, 0, 5/2)^{\mathrm{T}}$，最优值 $z^* = 20$。

值得注意的是，在最终单纯形表 2-9 中，除基变量的检验数为零外，非基变量 x_3 的检验数也为零，这意味着若让 x_3 改变取值不会使目标函数值有所改变。如果让 x_3 进基并继续迭代，就会得到另一个基本可行解，见表 2-10。

北京大学出版社版权所有 禁止转载

表 2-10 单纯形表

C_B	X_B	$B^{-1}b$	2	4	0	0	0	θ
			x_1	x_2	x_3	x_4	x_5	
4	x_2	7/2	0	1	1/4	1/4	0	14
2	x_1	3	1	0	−1/2	1/2	0	—
0	x_5	5/2	0	0	[3/4]	−1/4	1	10/3
检验数 σ			0	0	0	−2	0	
4	x_2	8/3	0	1	0	1/3	−1/3	
2	x_1	14/3	1	0	0	1/3	2/3	
0	x_3	10/3	0	0	1	−1/3	4/3	
检验数 σ			0	0	0	−2	0	

最优解 $\boldsymbol{X}_2^* = (14/3, 8/3, 10/3, 0, 0)^{\mathrm{T}}$，最优值为 $z^* = 20$。

根据最优解的定义，使目标函数达到最大值的任一可行解都是最优解，这样该线性规划有多个最优解，其最优解的一般表达式为

$$\boldsymbol{X}^* = \alpha \boldsymbol{X}_1^* + (1-\alpha)\boldsymbol{X}_2^*, \quad 0 \leqslant \alpha \leqslant 1$$

表 2-9 和表 2-10 求出的两个最优解是该线性规划可行域的两个顶点，据前面的讨论可知，这两个顶点连线上的一切点都是该问题的最优解，如点 $(23/6, 37/12)$，即 $x_1 = 23/6$，$x_2 = 37/12$，所对应的目标函数值也是 20，它虽不是基本可行解(不在可行域的顶点上)，但同样是该线性规划问题的一个最优解。

2.2.4 单纯形法的进一步讨论

单纯形法的求解基础是能够找到线性规划的可行解，如何找到这个可行解呢？这里学习构造初始可行解的大 M 法。

1. 大 M 法

在用单纯形法求解线性规划问题时，首先要找到一个初始可行基(一般是单位矩阵)。若其系数矩阵中已有一个单位矩阵，则可以作为初始可行基；若不存在单位矩阵，则可以采用在每个约束等式中人为地添加一个非负变量(称为人工变量)的方法，来构造另外一个线性规划问题，下面通过实例说明如何使用大 M 法。

【例 2.14】求解线性规划。

$$\max z = 3x_1 - x_2 - x_3$$

$$\begin{cases} x_1 - 2x_2 + x_3 \leqslant 11 \\ -4x_1 + x_2 + 2x_3 \geqslant 3 \\ -2x_1 + x_3 = 1 \\ x_1, \ x_2, \ x_3 \geqslant 0 \end{cases}$$

解：首先添加松弛变量 x_4 和剩余变量 x_5，将其化为标准型，得

$$\max z = 3x_1 - x_2 - x_3$$

$$\begin{cases} x_1 - 2x_2 + x_3 + x_4 = 11 \\ -4x_1 + x_2 + 2x_3 - x_5 = 3 \\ -2x_1 + x_3 = 1 \\ x_1, \ x_2, \ x_3, \ x_4, \ x_5 \geqslant 0 \end{cases} \tag{2-13}$$

它的系数矩阵为

$$\boldsymbol{A} = \begin{bmatrix} 1 & -2 & 1 & 1 & 0 \\ -4 & 1 & 2 & 0 & -1 \\ -2 & 0 & 1 & 0 & 0 \end{bmatrix}$$

矩阵 $\boldsymbol{A}$ 中有一个单位矩阵列，但不含有单位矩阵。在第二、三个约束方程的左边添加人工变量 $x_6, x_7 \geqslant 0$，就构成一个系数矩阵中含有单位矩阵的新的线性规划问题。

$$\max z = 3x_1 - x_2 - x_3$$

$$\begin{cases} x_1 - 2x_2 + x_3 + x_4 = 11 \\ -4x_1 + x_2 + 2x_3 - x_5 + x_6 = 3 \\ -2x_1 + x_3 + x_7 = 1 \\ x_1, \ x_2, \ x_3, \ x_4, \ x_5, \ x_6, \ x_7 \geqslant 0 \end{cases} \tag{2-14}$$

只有当所添加的人工变量取值为 0 时，线性规划(2-14)才和原规划(2-13)等价。为此，在目标函数中对人工变量的价值系数进行惩罚，使其在求解过程中取值为零，最好为非基变量，所以在求 max 问题中，人工变量在目标函数中的系数均为$-M$；在 min 问题中，人工变量在目标函数中的系数均为 M，这里 M 是一个任意大的正数。大 M 法也叫罚系数法，利用大 M 法的关键在于如何能够尽快将人工变量替换出基。

北京大学出版社版权所有 禁止转载

添加人工变量后构成的新的线性规划为

$$\max z = 3x_1 - x_2 - x_3 - 0x_4 - 0x_5 - Mx_6 - Mx_7$$

$$\begin{cases} x_1 - 2x_2 + x_3 + x_4 = 11 \\ -4x_1 + x_2 + 2x_3 - x_5 + x_6 = 3 \\ -2x_1 + x_3 + x_7 = 1 \\ x_1,\ x_2,\ x_3,\ x_4,\ x_5,\ x_6,\ x_7 \geqslant 0 \end{cases}$$

这样，就可以将人工变量视为基变量，得到一个初始基本可行解，就可用单纯形法进行迭代计算。若经若干次迭代，人工变量被全部替换出了基，则原问题的约束条件得到恢复，同时也有了一个基本可行解，原规划单纯形求解的初始条件就已具备，可继续采用单纯形表求解。

例 2.14 的求解过程见表 2-11。

北京大学出版社版权所有
禁止转载

表 2-11　单纯形表

C_B	X_B	$B^{-1}b$	3	−1	−1	0	0	$-M$	$-M$	θ
			x_1	x_2	x_3	x_4	x_5	x_6	x_7	
0	x_4	11	1	−2	1	1	0	0	0	11
$-M$	x_6	3	−4	1	2	0	−1	1	0	3/2
$-M$	x_7	1	−2	0	[1]	0	0	0	1	1
检验数 σ			$3-6M$	$-1+M$	$-1+3M$	0	$-M$	0	0	
0	x_4	10	3	−2	0	1	0	0	−1	—
$-M$	x_6	1	0	[1]	0	0	−1	1	−2	1
−1	x_3	1	−2	0	1	0	0	0	1	—
检验数 σ			1	$-1+M$	0	0	$-M$	0	$-3M+1$	
0	x_4	12	[3]	0	0	1	−2	2	−5	4
−1	x_2	1	0	1	0	0	−1	1	−2	—
−1	x_3	1	−2	0	1	0	0	0	1	—
检验数 σ			1	0	0	0	−1	$1-M$	$-1-M$	
3	x_1	4	1	0	0	1/3	−2/3	2/3	−5/3	
−1	x_2	1	0	1	0	0	−1	1	−2	—
−1	x_3	9	0	0	1	2/3	−4/3	4/3	−7/3	—
检验数 σ			0	0	0	−1/3	−1/3	$1/3-M$	$2/3-M$	

最优解 $(x_1, x_2, x_3, x_4, x_5, x_6, x_7)^{\mathrm{T}} = (4, 1, 9, 0, 0, 0, 0)^{\mathrm{T}}$，最优值 $z = 2$ 。

大 M 法求解线性规划会出现以下两种情况：①若最优解的基变量中含有人工变量，可以证明原问题无可行解；②若最优解的基变量中不含人工变量，即人工变量为非基变量，可以证明从最优解中去掉人工变量即为原问题的最优解。

【例 2.15】利用大 M 法求解线性规划。

$$\max z = -3x_1 + x_3$$

$$\begin{cases} x_1 + x_2 + x_3 \leqslant 4 \\ -2x_1 + x_2 - x_3 \geqslant 1 \\ 3x_2 + x_3 = 9 \\ x_1 \geqslant 0,\ x_2 \geqslant 0,\ x_3 \geqslant 0 \end{cases}$$

解：将该线性规划标准化，得

$$\max z = -3x_1 + x_3 + 0x_4 + 0x_5$$

$$\begin{cases} x_1 + x_2 + x_3 + x_4 = 4 \\ -2x_1 + x_2 - x_3 - x_5 = 1 \\ 3x_2 + x_3 = 9 \\ x_1,\ x_2,\ x_3,\ x_4,\ x_5 \geqslant 0 \end{cases}$$

添加人工变量，得

$$\max z = -3x_1 + x_3 + 0x_4 + 0x_5 - Mx_6 - Mx_7$$

$$\begin{cases} x_1 + x_2 + x_3 + x_4 = 4 \\ -2x_1 + x_2 - x_3 - x_5 + x_6 = 1 \\ 3x_2 + x_3 + x_7 = 9 \\ x_1,\ x_2,\ x_3,\ x_4,\ x_5,\ x_6,\ x_7 \geqslant 0 \end{cases}$$

北京大学出版社版权所有 禁止转载

列出初始单纯形表，进行单纯形迭代，见表 2-12(a)。

表 2-12(a)　单纯形表

C_B	X_B	$B^{-1}b$	−3	0	1	0	0	−M	−M	θ
			x_1	x_2	x_3	x_4	x_5	x_6	x_7	
0	x_4	4	1	1	1	1	0	0	0	4
−M	x_6	1	−2	**[1]**	−1	0	−1	1	0	1
−M	x_7	9	0	3	1	0	0	0	1	3
检验数 σ			−3−2M	4M	1	0	−M	0	0	
0	x_4	3	3	0	2	1	1	−1	0	1
0	x_2	1	−2	1	−1	0	−1	1	0	—
−M	x_7	6	**[6]**	0	4	0	3	−3	1	1
检验数 σ			6M−3	0	4M+1	0	3M	−4M	0	
0	x_4	0	0	0	0	1	−1/2	−1/2	1/2	—
0	x_2	3	0	1	1/3	0	0	0	1/3	9
−3	x_1	1	**1**	0	**[2/3]**	0	1/2	−1/2	1/6	3/2
检验数 σ			0	0	3	0	3/2	−M−3/2	−M+1/2	

从表 2-12(a)中看到人工变量已不在基变量中，可去掉人工变量进行迭代，得表 2-12(b)。

表 2-12(b)　单纯形表

C_B	X_B	$B^{-1}b$	−3	0	1	0	0			θ
			x_1	x_2	x_3	x_4	x_5			
0	x_4	0	0	0	0	1	−1/2			
0	x_2	5/2	−1/2	1	0	0	−1/4			
1	x_3	3/2	3/2	0	1	0	3/4			
检验数 σ			−9/2	0	0	0	−3/4			

由于所有的检验数为非正，得到最优解 $\boldsymbol{X}^* = (0,\ \ 5/2,\ \ 3/2,\ \ 0,\ \ 0,\ \ 0,\ \ 0)^{\mathrm{T}}$，最优值为 z^*=3/2。

2. 解的退化与循环

在单纯形法中，基变量一般都取非零值，非基变量都取零值。如果某个基本可行解中存在取零值的基变量，则称该解为退化解。退化解发生在单纯形换基迭代过程中，当利用最小比值原则确定换基变量时，若存在两个以上相同的最小比值，则在下一次迭代中就有一个或几个基变量为零。在退化情况下，如果取退化的基变量为出基变量，则变化后的解仍为退化解，且目标函数值不变。在以后的迭代中，如果每次都取退化的基变量为换出变量，则迭代可能只在可行域的几个顶点中间反复进行，即出现计算过程的循环，而达不到最优解。但实际中，循环现象极为罕见，计算的时候可以不考虑循环问题。

北京大学出版社版权所有　禁止转载

2.3　对偶问题和灵敏度分析

线性规划与其对偶问题是对同一问题的不同诠释。任何一个最大化的线性规划都有一个最小化的线性规划问题与之对应，称这一对互相联系的两个问题为一对对偶问题。本节将讨论线性规划的对偶问题和灵敏度分析，从而加深对线性规划问题的理解，扩大其应用范围。

2.3.1　线性规划的对偶问题

线性规划的原型是生产计划问题；对偶问题的原型是资源定价问题。

1. 对偶问题的提出

【例 2.16】格利公司计划制造Ⅰ、Ⅱ两种家电产品。已知制造一件分别需要的钢材、铜材和设备台时，每天可用于这两种家电的资源、出售一件时的获利情况见表 2-13。问该公司应制造两种家电各多少件，获利最大？

表 2-13　获利表

项　　目	Ⅰ	Ⅱ	资源数量
钢　　材	9	4	36
铜　　材	4	5	20
设备台时	3	10	30
利　　润	70	120	

解：这是一个生产计划问题，设生产这两种产品的数量为x_1, x_2，其线性规划数学模型为

$$\max z = 70x_1 + 120x_2$$
$$\begin{cases} 9x_1 + 4x_2 \leqslant 36 \\ 4x_1 + 5x_2 \leqslant 20 \\ 3x_1 + 10x_2 \leqslant 30 \\ x_1, \quad x_2 \geqslant 0 \end{cases} \tag{2-15}$$

现在，从另一个角度来考虑该问题，假设该企业愿出租、出让自己的资源，他想将自己生产产品(自营)改为对外加工(外包)，此时，决策者必须考虑如何为这三种资源定价的问题。

设 y_1, y_2, y_3 分别代表转让两种资源和出租设备的价格和租金。定价的原则是：生产一个单位的Ⅰ产品需消耗 9 个单位的钢材、4 个单位的铜材、3 个单位的设备台时，获利 70 个单位(打包参考)；那么，将这些资源全部转让时所获得的利润$9y_1 + 4y_2 + 3y_3$应不少于 70 个单位，即

$$9y_1 + 4y_2 + 3y_3 \geqslant 70 \tag{2-16}$$

同样地分析，有

$$4y_1 + 5y_2 + 10y_3 \geqslant 120 \tag{2-17}$$

此时，企业的总获利(即对方的总付出)为

$$w = 36y_1 + 20y_2 + 30y_3 \tag{2-18}$$

而对方一定希望总付出为最小，为使双方达成协议，该厂只能在约束条件(2-16)和(2-17) 下求式(2-18)的最小值，即

$$\min w = 36y_1 + 20y_2 + 30y_3$$
$$\begin{cases} 9y_1 + 4y_2 + 3y_3 \geqslant 70 \\ 4y_1 + 5y_2 + 10y_3 \geqslant 120 \\ y_1, \quad y_2, \quad y_3 \geqslant 0 \end{cases} \tag{2-19}$$

两个模型(2-15)和(2-19)是对同一问题两种不同决策的数学描述，它们之间有一定的内在联系。进行比较分析，得知两个模型具有以下对应关系。

(1) 两个问题的系数矩阵互为转置。

(2) 一个问题的变量个数等于另一个问题的约束条件个数。

(3) 一个问题的右端系数是另一个问题的目标函数的系数。

(4) 一个问题的目标函数为最大化，约束条件为“≤”类型；另一个问题的目标函数为最小化，约束条件为“≥”。

北京大学出版社版权所有 禁止转载

将满足上述条件的对应关系称为对偶关系，如果把式(2-15)线性规划称为原问题，则式 (2-19)称为对偶问题。一般地，原问题 LP 的对偶问题为 DP：

原问题 LP

$$\max z = c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

$$\begin{cases} a_{11}x_1 & +a_{12}x_2 & +\cdots & +a_{1n}x_n & \leqslant b_1 \\ a_{21}x_1 & +a_{22}x_2 & +\cdots & +a_{2n}x_n & \leqslant b_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1}x_1 & +a_{m2}x_2 & +\cdots & +a_{mn}x_n & \leqslant b_m \\ x_j \geqslant 0 & (j=1, & 2, & \cdots, & n) \end{cases}$$

对偶问题 DP

$$\min w = b_1y_1 + b_2y_2 + \cdots + b_my_m$$

$$\begin{cases} a_{11}y_1 & +a_{21}y_2 & +\cdots & +a_{m1}y_m & \geqslant c_1 \\ a_{12}y_1 & +a_{22}y_2 & +\cdots & +a_{m2}y_m & \geqslant c_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{1n}y_1 & +a_{2n}y_2 & +\cdots & +a_{mn}y_m & \geqslant c_n \\ y_i \geqslant 0 & (i=1, & 2, & \cdots, & m) \end{cases}$$

原问题和对偶问题之间的对应关系也可用矩阵表示，用矩阵符号表示原问题 LP 和对偶问题 DP 为

原问题 LP

$$\max z = \boldsymbol{CX} \qquad \begin{cases} \boldsymbol{AX} \leqslant \boldsymbol{b} \\ \boldsymbol{X} \geqslant \boldsymbol{0} \end{cases} \tag{2-20}$$

对偶问题 DP

$$\min w = \boldsymbol{Yb} \qquad \begin{cases} \boldsymbol{YA} \geqslant \boldsymbol{C} \\ \boldsymbol{Y} \geqslant \boldsymbol{0} \end{cases} \tag{2-21}$$

其中 $\boldsymbol{Y} = (y_1, y_2, \cdots, y_m)$ 是一个行向量，称为对偶向量，其余符号的含义已在 2.1.2 节中介绍。

2. 原问题与其对偶问题的关系

这里从一个具体的例子来说明线性规划与其对偶问题的关系。

【例 2.17】求下列原问题的对偶问题。

$$\max z = 5x_1 - 6x_2 + 7x_3 + 4x_4$$

$$\begin{cases} x_1 & +2x_2 & -x_3 & -x_4 & = -7 \\ 6x_1 & -3x_2 & +x_3 & -7x_4 & \leqslant 14 \\ -28x_1 & -17x_2 & +4x_3 & +2x_4 & \geqslant -3 \\ x_1, & x_2, & x_3, & x_4, & \geqslant 0 \end{cases}$$

解：为求原问题的对偶问题，首先将原问题转化为(2-20)的形式。

第一约束等式等价于下面两个不等式约束：

$$x_1 + 2x_2 - x_3 - x_4 \leqslant -7$$

$$x_1 + 2x_2 - x_3 - x_4 \geqslant -7 \qquad -x_1 - 2x_2 + x_3 + x_4 \leqslant 7$$

第二个约束不等式照写：

$$6x_1 - 3x_2 + x_3 - 7x_4 \leqslant 14$$

第三个不等式变成：

$$28x_1 + 17x_2 - 4x_3 - 2x_4 \leqslant 3$$

以 y_1^1, y_1^2, y_2, y_3 分别表示这 4 个不等式约束对应的对偶变量，则对偶问题为

$$\min w = -7y_1^1 + 7y_1^2 + 14y_2 + 3y_3$$

$$\begin{cases} y_1^1 & -y_1^2 & +6y_2 & +28y_3 & \geqslant 5 \\ 2y_1^1 & -2y_1^2 & -3y_2 & +17y_3 & \geqslant -6 \\ -y_1^1 & +y_1^2 & +y_2 & -4y_3 & \geqslant 7 \\ -y_1^1 & +y_1^2 & -7y_2 & -2y_3 & \geqslant 4 \\ y_1^1, & y_1^2, & y_2, & y_3 & \geqslant 0 \end{cases}$$

北京大学出版社版权所有 禁止转载

令 $y_1 = y_1^1 - y_1^2$，则上式的对偶问题变为

$$\min w = -7y_1 + 14y_2 + 3y_3$$

$$\begin{cases} y_1 + 6y_2 + 28y_3 \geqslant 5 \\ 2y_1 - 3y_2 + 17y_3 \geqslant -6 \\ -y_1 + y_2 - 4y_3 \geqslant 7 \\ -y_1 - 7y_2 - 2y_3 \geqslant 4 \\ y_1\text{无约束}, \quad y_2 \geqslant 0, \quad y_3 \geqslant 0 \end{cases}$$

同理可以导出，若原问题中的某个变量无非负限制，则对偶问题中的相应约束为等式。总之，可以把原问题和对偶问题之间的关系归纳成表 2-14 所给出的关系，就可直接写出其对偶问题。

表 2-14　原问题与对偶问题的关系

原问题(或对偶问题)	对偶问题(或原问题)
目标函数 max z	目标函数 min w
n 个变量	n 个约束
变量≥	约束条件≥
变量≤	约束条件≤
自由变量	约束条件=
m 个约束条件	m 个变量
约束条件≥	变量≤
约束条件≤	变量≥
约束条件=	自由变量
约束条件右端项	目标函数变量的系数
目标函数变量的系数	约束条件右端项

北京大学出版社版权所有 禁止转载

【例 2.18】写出下列线性规划的对偶问题。

$$\max z = x_1 + 2x_2 + x_3$$

$$\begin{cases} x_1 - x_2 + x_3 \leqslant 2 \\ -x_1 + x_2 + x_3 = 1 \\ 2x_1 + x_2 + x_3 \geqslant 2 \\ x_1 \geqslant 0, \quad x_2, \quad x_3\text{无约束} \end{cases}$$

解：根据表 2-14 给出的关系可直接写出该线性规划的对偶问题

$$\min w = 2y_1 + y_2 + 2y_3 \qquad\qquad \min w = 2y_1 + y_2 - 2y_3$$

$$\begin{cases} y_1 - y_2 + 2y_3 \geqslant 1 \\ -y_1 + y_2 + y_3 = 2 \\ y_1 + y_2 + y_3 = 1 \\ y_1 \geqslant 0, \quad y_2\text{无约束}, \quad y_3 \leqslant 0 \end{cases} \qquad \begin{cases} y_1 - y_2 - 2y_3 \geqslant 1 \\ -y_1 + y_2 - y_3 = 2 \\ y_1 + y_2 - y_3 = 1 \\ y_1 \geqslant 0, \quad y_2\text{无约束}, \quad y_3 \geqslant 0 \end{cases}$$

也可将 y_3 变为非负约束，不改变变量的记号，参看上面右边的规划。

【例 2.19】直接写出原问题的对偶问题。

原问题 LP

$$\min w = 5x_1 - 6x_2 + 7x_3 + 4x_4$$

$$\begin{cases} x_1 & +2x_2 & -x_3 & -x_4 & = -7 \\ 6x_1 & -3x_2 & +x_3 & -7x_4 & \geqslant 14 \\ -2x_1 & -7x_2 & +4x_3 & +2x_4 & \leqslant -3 \\ x_1 \geqslant 0, & x_2 \leqslant 0, & x_3, & x_4\text{无约束} & \end{cases}$$

对偶问题 DP

$$\max z = -7y_1 + 14y_2 - 3y_3$$

$$\begin{cases} y_1 & +6y_2 & -2y_3 & \leqslant 5 \\ 2y_1 & -3y_2 & -7y_3 & \geqslant -6 \\ -y_1 & +y_2 & +4y_3 & = 7 \\ -y_1 & -7y_2 & +2y_3 & = 4 \\ y_1\text{无约束}, & y_2 \geqslant 0, & y_3 \leqslant 0 & \end{cases}$$

原问题的对偶问题的形式不唯一，但本质是一致的。

3. 线性规划的对偶理论

以下给出线性规划对偶问题的几个定理，这里不作证明。

定理 2.7 (对称性)线性规划对偶问题的对偶问题就是原问题。

定理 2.8 (弱对偶定理)若 $\overline{\boldsymbol{X}}$ 是问题(2-20)的可行解，$\overline{\boldsymbol{Y}}$ 是问题(2-21)的可行解，则有 $\boldsymbol{C}\overline{\boldsymbol{X}} \leqslant \overline{\boldsymbol{Y}}\boldsymbol{b}$。

推论 1　若 $\overline{\boldsymbol{X}}$ 和 $\overline{\boldsymbol{Y}}$ 分别是问题(2-20)和(2-21)的可行解，则(2-21)的最小值不会小于 $\boldsymbol{C}\overline{\boldsymbol{X}}$ [即 $\boldsymbol{C}\overline{\boldsymbol{X}}$ 为对偶问题(2-21)的目标函数的一个下界]；而原问题(2-20)的最大值不会大于 $\overline{\boldsymbol{Y}}\boldsymbol{b}$ [即 $\overline{\boldsymbol{Y}}\boldsymbol{b}$ 为原问题(2-20)的目标函数的一个上界]。

推论 2　互为对偶的一对线性规划问题，如果其中一个有可行解，但目标函数无界(求最大的目标函数无上界，求最小的目标函数无下界)，则另一个必无可行解。

定理 2.9 若 $\boldsymbol{X}^*$ 是问题(2-20)的可行解，$\boldsymbol{Y}^*$ 是问题(2-21)的可行解，且 $\boldsymbol{CX}^* = \boldsymbol{Y}^*\boldsymbol{b}$，则 $\boldsymbol{X}^*$ 和 $\boldsymbol{Y}^*$ 分别是对应线性规划的最优解。

定理 2.10 (强对偶定理)若线性规划问题(2-20)和(2-21)之一有最优解，则另一问题也有最优解，并且两者的目标函数值相等。

事实上，原问题(2-20)及其对偶问题可标准化为

原问题 LP

$$\max z = \boldsymbol{CX} + \boldsymbol{0X}_{SL}$$

$$\begin{cases} \boldsymbol{AX} + \boldsymbol{EX}_{SL} = \boldsymbol{b} \\ \boldsymbol{X}, \boldsymbol{X}_{SL} \geqslant \boldsymbol{0} \end{cases} \tag{2-22}$$

$$\boldsymbol{X}_{SL} = (x_{n+1}, x_{n+2}, \cdots, x_{n+m})^{\mathrm{T}}$$

对偶问题 DP

$$\min w = \boldsymbol{Yb} + \boldsymbol{0Y}_{SU}$$

$$\begin{cases} \boldsymbol{YA} - \boldsymbol{Y}_{SU}\boldsymbol{E} = \mathbf{C} \\ \boldsymbol{Y}, \boldsymbol{Y}_{SU} \geqslant \boldsymbol{0} \end{cases} \tag{2-23}$$

$$\boldsymbol{Y}_{SU} = (y_{m+1}, y_{m+2}, \cdots, y_{m+n})$$

设 $(\overline{\boldsymbol{X}}, \overline{\boldsymbol{X}}_{SL})$ 为原问题(2-22)的一个基本可行解(不一定为最优解)，它所对应的基矩阵为 $\boldsymbol{B}$，决策变量 $\overline{\boldsymbol{X}}$ 和松弛变量 $\overline{\boldsymbol{X}}_{SL}$ 所对应的检验数分别为：$\boldsymbol{C} - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}$，$-\boldsymbol{C}_B\boldsymbol{B}^{-1}$ (不一定满足“$\leqslant 0$”条件)。

令 $\boldsymbol{Y} = \boldsymbol{C}_B\boldsymbol{B}^{-1}$，这时两组检验数分别为：$\boldsymbol{C} - \boldsymbol{YA}$，$-\boldsymbol{Y}$。

结合对偶问题(2-23)，这两组检验数可分别记为 $-\boldsymbol{Y}_{SU}\boldsymbol{E} = \boldsymbol{C} - \boldsymbol{YA}$，$-\boldsymbol{Y}$，对应关系见表 2-15。

北京大学出版社版权所有 禁止转载

表 2-15　原规划检验数与对偶问题解的关系

	$\boldsymbol{X}$	$\boldsymbol{X}_{SL}$
$\boldsymbol{X}_B$	$\boldsymbol{B}^{-1}\boldsymbol{A}$	$\boldsymbol{B}^{-1}$
检验数	$\boldsymbol{C}-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}$	$-\boldsymbol{C}_B\boldsymbol{B}^{-1}$
	$-\boldsymbol{Y}_{SU}\ \boldsymbol{E}$	$-\boldsymbol{Y}$

若 $(\overline{\boldsymbol{X}}^*, \overline{\boldsymbol{X}}^*_{SL})$ 是问题式(2-20)的最优解，不妨设它所对应的基矩阵仍为 $\boldsymbol{B}$，令 $\boldsymbol{Y}^*=\boldsymbol{C}_B\boldsymbol{B}^{-1}$，以后 $\boldsymbol{C}_B\boldsymbol{B}^{-1}$ 也称为单纯形因子，则必定所有检验数(包括基变量的检验数)都小于零，即

$$\boldsymbol{C}-\boldsymbol{Y}^*\boldsymbol{A},\quad -\boldsymbol{Y}^*$$

都小于零。

得

$$\boldsymbol{Y}^*\boldsymbol{A}\geqslant \boldsymbol{C},\quad \boldsymbol{Y}^*\geqslant 0$$

可见 $\boldsymbol{Y}^*$ 满足问题式(2-21)的约束条件，故 $\boldsymbol{Y}^*$ 是式(2-21)的可行解，其目标函数值为：$w=\boldsymbol{Y}^*\boldsymbol{b}=\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b}$，因问题式(2-20)的最优解为 X^*，它的目标函数值为 $z=\boldsymbol{C}\boldsymbol{X}^*=\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b}$，由此得到 $\boldsymbol{C}\boldsymbol{X}^*=\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b}=\boldsymbol{Y}^*\boldsymbol{b}$，即 $\boldsymbol{Y}^*$ 是式(2-21)的最优解。

强对偶定理及其导出过程给出了一个重要的结论：将原始单纯形表中松弛变量的检验数反号恰好得到对偶问题的一个解。由此得到下述重要结论。

(1) 在原问题的单纯形表中，原问题的松弛变量的检验数对应于对偶问题的决策变量，而原问题的决策变量的检验数对应于对偶问题的剩余变量，只是符号相反。

(2) 在获得最优解之前，$\boldsymbol{C}-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}$ 及 $-\boldsymbol{C}_B\boldsymbol{B}^{-1}$ 的各分量中至少有一大于零，即 $\boldsymbol{Y}_{SU}$ 和 $\boldsymbol{Y}$ 中至少有一个小于零，这时对应的对偶问题的解为非可行解。当原问题获得最优解时，表明 $\boldsymbol{C}-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}\leqslant 0$ 和 $-\boldsymbol{C}_B\boldsymbol{B}^{-1}\leqslant 0$，即 $\boldsymbol{Y}_{SU}\geqslant 0$，$\boldsymbol{Y}\geqslant 0$，此时对偶问题也获得最优解。

【例 2.20】求例 2.8 线性规划对偶问题的最优解。

原问题 *LP*

$$\max z=2x_1+x_2$$

$$\begin{cases}3x_1+5x_2\leqslant 15\\6x_1+2x_2\leqslant 24\\x_1,\ x_2\geqslant 0\end{cases}$$

对偶问题 *DP*

$$\min w=15y_1+24y_2$$

$$\begin{cases}3y_1+6y_2\geqslant 2\\5y_1+2y_2\geqslant 1\\y_1,\ y_2\geqslant 0\end{cases}$$

解：原问题的单纯形表见表 2-16。

表 2-16　单纯形表

$\boldsymbol{C}_B$	$\boldsymbol{X}_B$	$\boldsymbol{B}^{-1}\boldsymbol{b}$	2	1	0	0	θ
			x_1	x_2	x_3	x_4	
1	x_2	3/4	0	1	1/4	−1/8	
2	x_1	15/4	1	0	−1/12	5/24	
检验数 σ			0	0	−1/12	−7/24	

据上面的结论，其对偶问题的最优解可直接由最优单纯形表的检验数得出，即最优解为

$$\boldsymbol{Y}^*=(y_1,y_2)=(-\sigma_3,-\sigma_4)=(1/12,7/24)$$

北京大学出版社版权所有 禁止转载

代入对偶问题的目标函数得

$$w^* = 1/12 \times 15 + 7/24 \times 24 = 33/4$$

上面的分析表明：在两个互为对偶问题的线性规划问题中，可以任选一个求解，当然，通常选择约束条件少的一个求解。

定理 2.11 (松弛互补定理)　考虑对偶问题(2-20)和(2-21)，设 $\boldsymbol{X}^*$ 和 $\boldsymbol{Y}^*$ 分别是(LP)和(DP)的可行解，则它们同时也是最优解的充要条件为

$$\begin{cases}(\boldsymbol{C}-\boldsymbol{Y}^*\boldsymbol{A})\boldsymbol{X}^*=\boldsymbol{0}\\ \boldsymbol{Y}^*(\boldsymbol{A}\boldsymbol{X}^*-\boldsymbol{b})=\boldsymbol{0}\end{cases} \tag{2-24}$$

由此可得以下结论。

(1) 若式(2-20)有最优解 $\boldsymbol{X}^*$，使得对变量 $x_j > 0$ [称为 j 对式(2-20)是松的]，则对式(2-21)的一切最优解 $\boldsymbol{Y}^*$，必有 $\boldsymbol{Y}^*\boldsymbol{P}_j = c_j$，即对偶问题的第 j 个约束为等式[称为 j 对式(2-21)是紧的]。

(2) 若式(2-21)有最优解 $\boldsymbol{Y}^*$，使对约束满足 $\boldsymbol{Y}^*\boldsymbol{P}_j > c_j$ [称 j 对式(2-21)是松的]，则对(LP)的一切最优解 $\boldsymbol{X}^*$，必有 $x_j = 0$ [称 j 对式(2-20)是紧的]。

【例 2.21】给定一组对偶问题：

原问题 LP

$$\max z = 3x_1 + 2x_2 + 8x_3$$

$$\begin{cases}-4x_1 + 3x_2 - 12x_3 \leqslant -12\\ x_1 + 4x_3 \leqslant 6\\ x_2 - x_3 = 2\\ x_1,\ x_2,\ x_3 \geqslant 0\end{cases}$$

对偶问题 DP

$$\min w = -12y_1 + 6y_2 + 2y_3$$

$$\begin{cases}-4y_1 + y_2 \geqslant 3\\ 3y_1 + y_3 \geqslant 2\\ -12y_1 + 4y_2 - y_3 \geqslant 8\\ y_1 \geqslant 0,\ y_2 \geqslant 0,\ y_3\text{无约束}\end{cases}$$

若 $\boldsymbol{X}^* = (x_1, x_2, x_3)^{\mathrm{T}} = (6, 2, 0)^{\mathrm{T}}$ 为原问题的最优解，试用松弛互补定理求出对偶问题的最优解。

解：设 $\boldsymbol{Y}^* = (y_1, y_2, y_3)$ 是对偶问题的最优解，根据松弛互补定理，如果 $\boldsymbol{X}^* = (6, 2, 0)^{\mathrm{T}}$ 是原问题的最优解，那么 $\boldsymbol{Y}^*$ 除了应是对偶问题的一个可行解外，还应满足松弛互补条件。

因为 $x_1 = 6 > 0$，所以 $-4y_1 + y_2 = 3$；同理可知 $3y_1 + y_3 = 2$；又因为 $-4x_1 + 3x_2 - 12x_3 = -18 < -12$，所以 $y_1 = 0$，于是可求得 $\boldsymbol{Y}^* = (y_1, y_2, y_3) = (0, 3, 2)$，不难验证它是对偶问题的可行解，且又满足松弛互补条件，因此它是对偶问题的最优解，最优值 $z^* = w^* = 22$。

4. 对偶问题的经济解释

对于线性规划问题(2-20)，当用它来处理资源分配问题时，其决策变量代表的是产品的产量，它的对偶问题(2-21)就是资源定价问题，其对偶变量也有明显的经济意义。

事实上，若 $\boldsymbol{X}^*$ 为原问题(2-20)的最优解，最优目标函数值为 z^*，根据对偶定理，对偶问题也有最优解 $\boldsymbol{Y}^*$，且两者的最优目标函数值相等，即

$$z^* = \sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^* = w^*$$

这就是说，原问题的目标函数值也可以看成是各个 $b_i y_i^*$ 相加而成的，故可将每个 $b_i y_i^*$ 看成是第 i 种资源对目标函数值所做的贡献，由于 b_i 是第 i 种资源的拥有量，因此，$b_i y_i^* / b_i = y_i^*$ 便可以理解为是每个单位的第 i 种资源对目标函数值的贡献，即增加或减少单位第 i 种资源所引起总收益(目标函数)的改变量，称 y_i^* 为第 i 种资源的影子价格。

北京大学出版社版权所有 禁止转载

显然这种价格不同于第 i 种资源的市场价格，它完全由企业内部的生产条件(不是由市场)所决定，是企业内部决策的一种参照价格，便利于企业更好地控制资源的使用。同一种资源在不同的企业影子价格一般可以不同，一种资源的影子价格越大，则增加或减少一个单位这种资源，对总收益的影响越大；如果一种资源的影子价格为零，则在一定范围内增加或减少一个单位这种资源对总收益没有影响。

【例 2.22】某公司隶属的甲、乙两个工厂有 A 原料 360 千克，B 原料 640 千克。甲厂用 A、B 两种原料生产Ⅰ、Ⅱ两种产品，乙厂用 A、B 两种原料生产Ⅲ、Ⅳ两种产品。每种单位产品所消耗各种原料的数量及利润、原料分配等见表 2-17。

表 2-17　利润、原料分配表

工　厂		甲			乙		
		Ⅰ	Ⅱ	分配原料	Ⅲ	Ⅳ	分配原料
原料/千克	A	8	4	160	5	8	200
	B	6	10	330	10	4	310
利润/万元		4	3		3	4	

解：建立两工厂的线性规划模型为

甲厂的数学模型为

$$\max z_1 = 4x_1 + 3x_2$$

$$\begin{cases} 8x_1 + 4x_2 \leqslant 160 \\ 6x_1 + 10x_2 \leqslant 330 \\ x_1, \quad x_2 \geqslant 0 \end{cases}$$

乙厂的数学模型为

$$\max z_2 = 3x_3 + 4x_4$$

$$\begin{cases} 5x_3 + 8x_4 \leqslant 200 \\ 10x_3 + 4x_4 \leqslant 310 \\ x_3, \quad x_4 \geqslant 0 \end{cases}$$

数模的最优单纯形表分别列在表 2-18 的左、右两边，该公司的总产值为 $z = z_1 + z_2 = 110 + 114 = 224$。

表 2-18　最优单纯形表

X_B	$B^{-1}b$	x_1	x_2	x_5	x_6	X_B	$B^{-1}b$	x_3	x_4	x_7	x_8
x_1	5	1	0	5/28	−1/14	x_3	28	1	0	−1/15	2/15
x_2	30	0	1	−3/28	1/7	x_4	15/2	0	1	1/6	−1/12
检验数 σ		0	0	−11/28	−1/7	检验数 σ		0	0	−7/15	−1/15

现在考虑公司能否制定出新的资源分配方案，使得利润最高。由表 2-17 可见，甲厂两种资源 A、B 的影子价格分别为 y_1=11/28，y_2=1/7；乙厂两种资源 A、B 的影子价格分别为 y_3=7/15，y_4=1/15。由于 $y_1 < y_3$，若减少甲厂 1 个单位的 A 种原料，而增加乙厂 1 个单位的 A 种原料，会增加总产值 $y_3 - y_1 \approx 0.07$ 万元；同理由于 $y_2 > y_4$，若减少乙厂 B 原料 1 kg，而增加甲厂 B 原料 1 kg，也会增加 $y_2 - y_4 \approx 0.08$ 万元。

如果甲厂减少 A 原料 1 kg 给乙厂，而乙厂减少 B 原料 1 kg 给甲厂，此时公司总产值能增加 $(0.07 \times 1 + 0.08 \times 1)$万元 $= 0.15$ 万元。

线性规划问题求最优解就是在有限资源条件下谋求最高的利益，此时相应的对偶问题中的变量就是影子价格。由于影子价格是指资源增加时对最优收益发生的影响，反映企业的资源利用效益，所以有人把它称为资源的边际产出或者资源的机会成本，它表示出资源在最优产品组合时，所具有的“潜在价值”或“贡献”。

北京大学出版社版权所有 禁止转载

影子价格在经营管理中的用处很多，一般来说可以提供以下几个方面的信息。

(1) 影子价格可以告诉管理人员，增加哪一种资源对增加经济效益更有利。例 2.22 甲厂两种资源的影子价格为(11/28, 1/7)，说明首先应考虑增加原料 A，因为相比之下它能使收益的增量更大，对企业决策来说，A 资源是更要关注的资源(核心资源)。

(2) 影子价格可以告诉管理人员，花多大的代价来增加资源才是合算的。在例 2.22 甲厂中，原料 A 每增加一个单位能使收益增加 11/28，如果增加这种资源的代价大于 11/28 就不合算了，这时企业决策应当考虑卖掉这种资源；反之企业应当考虑购买这种资源。同理，如果对偶问题的最优值大于原问题的最优值，企业应采用的决策是外包，反之采用的决策为自营。

(3) 影子价格可以告诉管理人员如何考虑新产品的价格。在例 2.22 甲厂中，企业要生产一种新产品，如果每件新产品耗用的这两种资源的数量是(7, 14)单位，则新产品的价格一定要大于 $(7\quad 14)(11/28\quad 1/7)^{\mathrm{T}}=\dfrac{11+8}{4}=4.75$ 才能增加收益，如价格低于 4.75 就不合适了。

(4) 影子价格可以告诉管理人员，当产品利润变动时哪些资源最为宝贵，哪些无关紧要。例 2.22 甲厂中，产品的利润由(4，3)变为(4，6)，则从单纯形表可以算出影子价格将从(11/28, 1/7)变为

$$\boldsymbol{C}_B\boldsymbol{B}^{-1}=[4\quad 6]\begin{bmatrix}5/28 & -1/14\\ -3/28 & 1/7\end{bmatrix}=[1/14\quad 4/7]$$

这说明第二种产品利润增加，原料 B 就显得更加宝贵了。

北京大学出版社版权所有 禁止转载

2.3.2 对偶单纯形法

1954 年美国数学家 C. 莱姆基提出对偶单纯形法。对偶单纯形法则是从满足对偶可行性条件出发通过迭代逐步搜索原问题的最优解，在迭代过程中始终保持基解的对偶可行性，而使不可行性逐步消失。所谓满足对偶可行性，即指其检验数满足最优性条件，因此在保持对偶可行性的前提下，一旦基解成为可行解时，也就得到最优解。

1. 对偶单纯形法的引入

【例 2.23】求解线性规划问题

$$\min w=15y_1+5y_2+11y_3$$
$$\begin{cases}3y_1+2y_2+2y_3\geqslant 5\\ 5y_1+y_2+2y_3\geqslant 4\\ y_1,\ y_2,\ y_3\geqslant 0\end{cases}$$

解：添加剩余变量，将上述问题转化为标准型，并将每个等式两边乘以-1，得

$$\min w=15y_1+5y_2+11y_3$$
$$\begin{cases}3y_1+2y_2+2y_3-y_4=5\\ 5y_1+y_2+2y_3-y_5=4\\ y_1,\ y_2,\ y_3,\ y_4,\ y_5\geqslant 0\end{cases}$$

$$\max(-w)=-15y_1-5y_2-11y_3$$
$$\begin{cases}-3y_1-2y_2-2y_3+y_4=-5\\ -5y_1-y_2-2y_3+y_5=-4\\ y_1,\ y_2,\ y_3,\ y_4,\ y_5\geqslant 0\end{cases}$$

如果取 $\boldsymbol{Y}_B=(y_4,y_5)$ 作为初始基变量，得到以下初始单纯形表，见表 2-19。

表 2-19　单纯形表

$\boldsymbol{C}_B$	$\boldsymbol{X}_B$	$\boldsymbol{B}^{-1}\boldsymbol{b}$	−15 y_1	−5 y_2	−11 y_3	0 y_4	0 y_5
0	y_4	−5	−3	**[−2]**	−2	1	0
0	y_5	−4	−5	−1	−2	0	1
	检验数 σ		−15	−5	−11	0	0

可以看出，两个基变量 $\boldsymbol{Y}_B=(y_4,y_5)$ 均取负值，所以 $\boldsymbol{Y}_B$ 所确定的基本解不是基本可行解，从而不能用单纯形法求解。但该例的所有检验数都小于 0，满足对偶可行性条件，可以利用对偶理论，仿照单纯形法的思想求解，这就是对偶单纯形法的来源。

此时 B 不是可行基，为保证上述方法的实现，可按下面的方法确定出基变量和进基变量。

确定出基变量：取一个具有最小负值的基变量为出基变量。在例 2.23 中，两个基变量 (y_4,y_5) 都取负值，且 $y_4=-5$ 最小，故 y_4 为出基变量。

确定进基变量：考虑出基变量 y_4 所对应的行中的所有负元素，对每个这样的元素作比值，根据最小比值原则：

$$\theta=\min\left\{\frac{-15}{-3},\frac{-5}{-2},\frac{-11}{-2}\right\}=\frac{5}{2}=\frac{\sigma_2}{(\boldsymbol{B}^{-1}\boldsymbol{P}_2)_1}$$

则 y_2 为进基变量。

由于 y_2 为进基变量，主元为 $a_{12}=-2$，对表 2-19 进行迭代便得表 2-20(a)。在表 2-20(a)中，基变量 $y_5=-3/2$，故 y_5 为出基变量；根据最小比值原则：

$$\theta=\min\left\{\frac{-15/2}{-7/2},\frac{-6}{-1},\frac{-5/2}{-1/2}\right\}=\frac{15}{7}=\frac{\sigma_1}{(\boldsymbol{B}^{-1}\boldsymbol{P}_1)_2}$$

得 y_1 是进基变量，主元为 $a_{21}=-7/2$，再作一次迭代，得表 2-20(b)。由于它的“左边”列全部非负，故它就是最终单纯形表。

最优解为 $\boldsymbol{Y}^*=(y_1,y_2,y_3,y_4,y_5)=(3/7,13/7,0,0,0)$，最优值 $w^*=110/7$。

表 2-20　单纯形表

$\boldsymbol{C}_B$	$\boldsymbol{X}_B$	$\boldsymbol{B}^{-1}\boldsymbol{b}$	−15 y_1	−5 y_2	−11 y_3	0 y_4	0 y_5	
−5	y_2	5/2	3/2	1	1	−1/2	0	(a)
0	y_5	−3/2	**[−7/2]**	0	−1	−1/2	1	
	检验数 σ		−15/2	0	−6	−5/2	0	
−5	y_2	13/7	0	1	4/7	−5/7	3/7	(b)
−15	y_1	3/7	1	0	2/7	1/7	−2/7	
	检验数 σ		0	0	−27/7	−10/7	−15/7	

2. 对偶单纯形法的计算步骤

对偶单纯形法的求解思想：在保证检验数行全部非正的条件下，逐步使得基解“左边”

北京大学出版社版权所有 禁止转载

一列各数变成非负，一旦“左边”一列各数均满足了非负条件(即可行性条件)，则就获得最优解。

从对偶单纯形法的思想出发，其计算步骤总结如下，如图 2.8 所示。

(1) 找出一个初始基 $\boldsymbol{B}_0$，要求对应的单纯形表中的全部检验数 $\sigma_j \leqslant 0$，但“左边”列中允许有负数，即确定出对偶可行解。

(2) 若“左边”列中各数均非负，则 $\boldsymbol{B}_0$ 已是最优基，于是求得最优解，计算终止；否则转为第(3)步。

(3) 若某个基变量取负值，但其所在的行中元素全部是正数，这时问题无可行解；否则转为第(4)步。

(4) 换基迭代：“左边”列中取值最小(即负得最多)的数所对应的变量为出基变量。为决定进基变量，必须按最小比值原则计算最小比值 θ，最小比值出现在哪一列，则该列所对应的变量即为进基变量，换基后得新基 $\boldsymbol{B}_1$，以出基变量的行和进基变量列交点处的元素为主元进行单纯形迭代，再转入第(2)步。

至于在一般情况下如何找出步骤(1)中的初始基 $\boldsymbol{B}_0$，这里不作讨论。

北京大学出版社版权所有
禁止转载

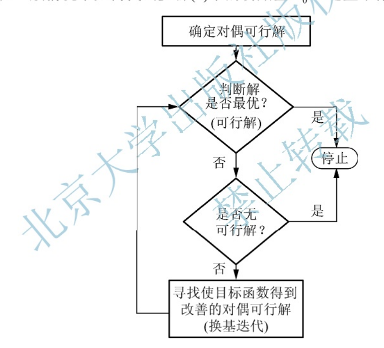


图 2.8 对偶单纯形法计算步骤

2.3.3 灵敏度分析

线性规划模型的确定是以 a_{ij}、b_i、c_j 为已知常数作为基础的，但在实际问题中，这些数据本身不仅很难准确得到，而且往往还要受到诸如市场价格波动，资源供应量变化，企业技术改造的因素的影响，因此，很自然地要提出这样的问题，当这些数据有一个或多个发生变化时，对已找到的最优解或最优基会产生怎样的影响；或者说这些数据在什么范围内变化，已找到的最优解或最优基不变；以及在原最优解或最优基不再是最优基时，如何用最简单的方法求出新的最优解或最优基。灵敏度分析要研究的就是这些问题。

【例 2.24】某工厂用甲、乙、丙 3 种原材料可生产 3 种产品，其中有关数据见表 2-21。

表 2-21　单纯形表

原　　料	供应量/千克	每件产品所需原料数/千克		
		A	B	C
甲	60	3	4	2
乙	40	2	1	2
丙	80	1	3	2
每件产品利润/万元		2	4	3

问怎样组织生产可以使工厂获得最大利润？

解：设 x_1、x_2、x_3 分别表示这 3 种产品的件数(可以为非成品)，建立线性规划模型为

$$\max z = 2x_1 + 4x_2 + 3x_3$$

$$\begin{cases} 3x_1 + 4x_2 + 2x_3 \leqslant 60 \\ 2x_1 + x_2 + 2x_3 \leqslant 40 \\ x_1 + 3x_2 + 2x_3 \leqslant 80 \\ x_1, x_2, x_3 \geqslant 0 \end{cases}$$

在上述各约束条件中依次加入松弛变量 x_4, x_5, x_6 转化为标准型，运用单纯形法求解上述模型，其运算过程见表 2-22，最后得到最优解为

$$X^* = (x_1, x_2, x_3, x_4, x_5, x_6)^{\mathrm{T}} = (0, 20/3, 50/3, 0, 0, 80/3)^{\mathrm{T}}$$

即最优生产方案是生产 B 产品 20/3 件，C 产品 50/3 件，最大利润 230/3 万元。

表 2-22　单纯形表

C_B	X_B	$B^{-1}b$	2	4	3	0	0	0	θ
			x_1	x_2	x_3	x_4	x_5	x_6	
4	x_2	20/3	1/3	1	0	1/3	−1/3	0	
3	x_3	50/3	5/6	0	1	−1/6	2/3	0	
0	x_6	80/3	−5/3	0	0	−2/3	−1/3	1	
检验数 σ			−11/6	0	0	−5/6	−2/3	0	

1. 价值向量的灵敏度分析

在线性规划的求解过程中，目标函数系数的变动将会影响检验数的取值，但是，当目标函数的系数的变动不破坏最优判别准则时，原最优解不变；否则，原最优解将发生变化，要设法求出新的最优解。下面分两种情况讨论。

1) c_j 是非基变量 x_j 的系数

在最终单纯形表上，x_j 所对应的检验数为 $\sigma_j = c_j - C_B B^{-1} P_j$。由于 c_j 是非基变量的系数，因此，它的改变对 $C_B B^{-1} P_j$ 取值不产生影响，而只影响 c_j 本身。

若 c_j 有一个增量 Δc_j，则变化后的检验数为

$$\sigma_j' = c_j + \Delta c_j - C_B B^{-1} P_j = \sigma_j + \Delta c_j$$

为保证原所求的解仍然为最优解，则要求检验数 σ_j' 仍满足最优解判别定理，故有

北京大学出版社版权所有 禁止转载

$$\sigma_j' = \sigma_j + \Delta c_j \leqslant 0 \quad \sigma_j \leqslant -\Delta c_j \tag{2-25}$$

2) c_k 是基变量 x_k 的系数

由于 c_k 是基变量的系数，则 c_k 是向量 $\boldsymbol{C}_B$ 的一个分量，当 c_k 改变量为 Δc_k 时，就引起 $\boldsymbol{C}_B$ 改变 $\Delta \boldsymbol{C}_B$，从而引起原问题最终表中的全体非基变量的检验数和目标函数值的改变，发生变化后的非基变量的检验数为

$$\begin{aligned}\sigma_j' &= c_j - (\boldsymbol{C}_B + \Delta \boldsymbol{C}_B)\boldsymbol{B}^{-1}\boldsymbol{P}_j \\ &= c_j - \boldsymbol{C}_B \boldsymbol{B}^{-1}\boldsymbol{P}_j - \Delta \boldsymbol{C}_B \boldsymbol{B}^{-1}\boldsymbol{P}_j \\ &= \sigma_j - \Delta c_k a_{kj}'\end{aligned}$$

其中

$$\Delta \boldsymbol{C}_B = (0, \cdots, 0, \Delta c_k, 0, \cdots, 0)$$

$$\boldsymbol{B}^{-1}\boldsymbol{P}_j = (a_{1j}', \cdots, a_{k-1j}', a_{kj}', a_{k+1j}', \cdots, a_{mj}')^{\mathrm{T}}$$

$\boldsymbol{P}_j$ 为 x_j 的系数列向量，a_{kj}' 是 $\boldsymbol{B}^{-1}\boldsymbol{P}_j$ 的第 k 个分量。

为保证原所求的解仍为最优解，则要求所有新的非基变量的检验数 σ_j' 仍满足最优判别条件，即有

$$\sigma_j' = \sigma_j - \Delta c_k a_{kj}' \leqslant 0$$

若 $a_{kj}' < 0$，则 $\Delta c_k \leqslant \dfrac{\sigma_j}{a_{kj}'}$；若 $a_{kj}' > 0$，则 $\Delta c_k \geqslant \dfrac{\sigma_j}{a_{kj}'}$，可得

$$\max\left\{\frac{\sigma_j}{a_{kj}'} \middle| a_{kj}' > 0\right\} \leqslant \Delta c_k \leqslant \min\left\{\frac{\sigma_j}{a_{kj}'} \middle| a_{kj}' < 0\right\} \tag{2-26}$$

【例 2.25】在例 2.24 中，考虑价值系数变化的影响。(1) c_1 在什么范围内变化时，最优基不变？(2)若 c_1 变为 4，是否对最优解产生影响？(3) c_2 在什么范围内变化时，最优基不变？(4)若 c_2 变为 9，求最优解。

解：(1) 根据公式(2-25)，可找到最优基的不变范围。

只要 $\sigma_1 = -11/6 \leqslant -\Delta c_1$，即 $c_1 = 2 + \Delta c_1 \leqslant 2 + 11/6 = 23/6$ 时，最优基就保持不变，所求出的最优解仍为最优解。

(2) 若 c_1 变为 4，由于 $4 > 23/6$，c_1 的变化会对最优解产生影响。

(3) 由于 $a_{11}' = 1/3$，$a_{14}' = 1/3$，$a_{15}' = -1/3$，所以：

$\sigma_1' = -11/6 - \Delta c_2 \times 1/3 \leqslant 0$，$\sigma_4' = -5/6 - \Delta c_2 \times 1/3 \leqslant 0$，$\sigma_5' = -2/3 + \Delta c_2 \times 1/3 \leqslant 0$，解上面不等式组得到 $-\dfrac{5}{2} \leqslant \Delta c_2 \leqslant 2$。

(4) 由题知 $\Delta c_2 = 5$，根据(3)知最优解发生改变。重新计算单纯形表 2-22 的检验数，其过程见表 2-23，得到最优解为

$$\boldsymbol{X}^* = (x_1, x_2, x_3, x_4, x_5, x_6)^{\mathrm{T}} = (0, 15, 0, 0, 25, 35)^{\mathrm{T}}$$

表 2-23　单纯形表

$\boldsymbol{C}_B$	$\boldsymbol{X}_B$	$\boldsymbol{B}^{-1}\boldsymbol{b}$	2	9	3	0	0	0	θ
			x_1	x_2	x_3	x_4	x_5	x_6	
9	x_2	20/3	1/3	1	0	1/3	−1/3	0	—
3	x_3	50/3	5/6	0	1	−1/6	[2/3]	0	25

北京大学出版社版权所有 禁止转载

续表

C_B	X_B	$B^{-1}b$	2	9	3	0	0	0	θ
			x_1	x_2	x_3	x_4	x_5	x_6	
0	x_6	80/3	−5/3	0	0	−2/3	−1/3	1	—
检验数 σ			−7/2	0	0	−5/2	1	0	
9	x_2	15	3/4	1	1/2	1/4	0	0	
0	x_5	25	5/4	0	3/2	−1/4	1	0	
0	x_6	35	−5/4	0	1/2	−3/4	0	1	
检验数 σ			−19/4	0	−3/2	−9/4	0	0	

2. 资源向量的灵敏度分析

资源项 b_r 的变化与最优判别条件无关，但它的变化将影响最终单纯形表中 $\boldsymbol{X}_B=\boldsymbol{B}^{-1}\boldsymbol{b}$ 的可行性。设变化后的 b' 仍能保证 $\boldsymbol{B}^{-1}\boldsymbol{b}'\geqslant 0$，这时 B 仍为最优基，$\boldsymbol{B}^{-1}\boldsymbol{b}'$ 为新的最优解，否则最优基将发生变化，此时需用对偶单纯形法继续迭代，直至求出最优解。下面研究 b_r 的变化范围。

设 b_r 有一个改变量 Δb_r，这时新的基本解为

$$\overline{\boldsymbol{X}}_B=\boldsymbol{B}^{-1}\begin{bmatrix} b_1 \\ \vdots \\ b_r+\Delta b_r \\ \vdots \\ b_m \end{bmatrix}=\boldsymbol{B}^{-1}\boldsymbol{b}+\boldsymbol{B}^{-1}\begin{bmatrix} 0 \\ \vdots \\ \Delta b_r \\ \vdots \\ 0 \end{bmatrix}$$

令 $\boldsymbol{\beta}_r=\begin{bmatrix}\beta_{1r} & \beta_{2r} \cdots & \beta_{mr}\end{bmatrix}^{\mathrm{T}}$ 为 $\boldsymbol{B}^{-1}$ 中的第 r 列，由于 $\boldsymbol{B}^{-1}\boldsymbol{b}$ 就是原来的基本可行解 $\boldsymbol{X}_B$，就有

$$\overline{\boldsymbol{X}}_B=\begin{bmatrix} b_1' \\ b_2' \\ \vdots \\ b_m' \end{bmatrix}+\Delta b_r\times\begin{bmatrix} \beta_{1r} \\ \beta_{2r} \\ \vdots \\ \beta_{mr} \end{bmatrix}$$

为了保持解的可行性，应有 $\overline{\boldsymbol{X}}_B\geqslant 0$，即

$$b_i'+\Delta b_r\beta_{ir}\geqslant 0,\ i=1,2,\cdots,m$$

若 $\beta_{ir}<0$，则 $\Delta b_r\leqslant\dfrac{-b'}{\beta_{ir}}$；若 $\beta_{ir}>0$，则 $\Delta b_r\geqslant\dfrac{-b_i'}{\beta_{ir}}$。

于是得到

$$\max\left\{\frac{-b_i'}{\beta_{ir}}\middle|\beta_{ir}>0\right\}\leqslant\Delta b_r\leqslant\min\left\{\frac{-b_i'}{\beta_{ir}}\middle|\beta_{ir}<0\right\} \tag{2-27}$$

【例 2.26】在例 2.24 中，考虑资源向量的变化。(1) b_1 在什么范围内变化，最优基保持不变？(2)求 b_2=90 时的最优解。

北京大学出版社版权所有 禁止转载

解：(1) 由表 2-22 可知，

$$B^{-1}=\begin{bmatrix}1/3 & -1/3 & 0\\ -1/6 & 2/3 & 0\\ -2/3 & -1/3 & 1\end{bmatrix} \qquad \begin{bmatrix}b_1'\\ b_2'\\ b_3'\end{bmatrix}=\begin{bmatrix}20/3\\ 50/3\\ 80/3\end{bmatrix}$$

如果 b_1 变化了 Δb_1，则根据式(2-27)有

$$\max\left\{-\frac{20/3}{1/3}\right\}\leqslant \Delta b_1 \leqslant \min\left\{\frac{50/3}{1/6},\frac{80/3}{2/3}\right\}$$

$$-20\leqslant \Delta b_1 \leqslant 40$$

由此可知 b_1 的变化范围为

$$40\leqslant b_1 \leqslant 100$$

(2) 当 b_2=90 时，由于

$$B^{-1}b'=\begin{bmatrix}1/3 & -1/3 & 0\\ -1/6 & 2/3 & 0\\ -2/3 & -1/3 & 1\end{bmatrix}\begin{bmatrix}60\\ 90\\ 80\end{bmatrix}=\begin{bmatrix}-10\\ 50\\ 10\end{bmatrix}$$

所以 b_2 改变后基解已不再是可行解，需继续求新的最优解，见表 2-24。

表 2-24 单纯形表

C_B	X_B	$B^{-1}b$	2	9	3	0	0	0
			x_1	x_2	x_3	x_4	x_5	x_6
4	x_2	−10	1/3	1	0	1/3	**[−1/3]**	0
3	x_3	50	5/6	0	1	−1/6	2/3	0
0	x_6	10	−5/3	0	0	−2/3	−1/3	1
	检验数 σ		−11/6	0	0	−5/6	−2/3	0
0	x_5	30	−1	−3	0	−1	1	0
3	x_3	30	−3/2	2	1	1/2	0	0
0	x_6	20	−2	−1	0	−1	0	1
	检验数 σ		−5/2	−2	0	−3/2	0	0

3. 追加一个新的决策变量

设增加一个新变量 x_{n+1}，其价值系数 c_{n+1}，在约束条件中对应的列向量为 $\boldsymbol{P}_{n+1}$，若 x_{n+1}=0，显然原问题的最优基是新问题的可行基，原有变量的检验数并没有改变，所以，可以将 x_{n+1} 看成非基变量，在原问题的最优单纯行表中增加一列：

$$\boldsymbol{P}'_{n+1}=\boldsymbol{B}^{-1}\boldsymbol{P}_{n+1} \qquad \sigma_{n+1}=c_{n+1}-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{P}_{n+1}$$

如果 $\sigma_{n+1}\leqslant 0$，则原问题的最优解就是新问题的最优解，否则继续迭代。

【例 2.27】在例 2.24 中，增加一个新变量 x_7，$c_7=4$，$\boldsymbol{P}_7=(2 \quad 1 \quad 4)^{\mathrm{T}}$，求增加新变量后线性规划的最优解。

北京大学出版社版权所有 禁止转载

解：在原可行基下计算其检验数

$$B^{-1}P_7=\begin{bmatrix}1/3 & -1/3 & 0\\ -1/6 & 2/3 & 0\\ -2/3 & -1/3 & 1\end{bmatrix}\begin{bmatrix}2\\1\\4\end{bmatrix}=\begin{bmatrix}1/3\\1/3\\7/3\end{bmatrix}$$

$$\sigma_7=c_7-C_BB^{-1}P_7=4-\begin{bmatrix}4 & 3 & 0\end{bmatrix}\begin{bmatrix}1/3\\1/3\\7/3\end{bmatrix}=\frac{5}{3}$$

由于 $\sigma_7>0$，需重新计算增加变量后线性规划的最优解，见表 2-25，得到追加变量后的最优解为 $X^*=(x_1,x_2,x_3,x_4,x_5,x_6,x_7)^{\mathrm{T}}=(0,20/7,90/7,0,0,0,80/7)^{\mathrm{T}}$ 。

表 2-25　单纯形表

C_B	X_B	$B^{-1}b$	2	4	3	0	0	0	4	θ
			x_1	x_2	x_3	x_4	x_5	x_6	x_7	
4	x_2	20/3	1/3	1	0	1/3	−1/3	0	1/3	20
3	x_3	50/3	5/6	0	1	−1/6	2/3	0	1/3	50
0	x_6	80/3	−5/3	0	0	−2/3	−1/3	1	[7/3]	80/7
检验数 σ			−11/6	0	0	−5/6	−2/3	0	5/3	
4	x_2	20/7	4/7	1	0	3/7	−2/7	−1/7	0	
3	x_3	90/7	15/14	0	1	−1/14	5/7	−1/7	0	
4	x_7	80/7	−5/7	0	0	−2/7	−1/7	3/7	1	
检验数 σ			−9/14	0	0	−5/14	−3/7	−5/7	0	

4. 技术矩阵 A 的灵敏度分析

(1) 非基变量 x_j 的系数变化时，其变化影响其检验数，即

$$\sigma_j'=c_j-C_BB^{-1}(P_j+\Delta P_j)=\sigma_j-C_BB^{-1}\Delta P_j\leqslant 0$$

则最优解与最优值均不变；若 $\sigma_j>0$，则需将原最优单纯形表中 x_j 对应的列修改为

$$\begin{aligned}&B^{-1}P_j\to B^{-1}(P_j+\Delta P_j)\\&c_j-C_BB^{-1}P_j\to c_j-C_BB^{-1}(P_j+\Delta P_j)\end{aligned}\tag{2-28}$$

然后用单纯形法迭代求解即可。

(2) 当基变量 x_j 的系数发生变化时，既影响到基本可行解，又影响到非基变量的检验数，仍按式(2-28)修改 x_j 对应的列，用单纯形和对偶单纯形法综合求出其最优解即可。

2.4　线性整数规划

整数线性规划与线性规划有着密不可分的关系，整数线性规划(Integer Linear Programming，ILP)问题研究的是要求变量取整数值时，在一组线性约束条件下某个线性函数的最优问题，是应用非常广泛的运筹学的重要分支。

北京大学出版社版权所有 禁止转载

2.4.1　线性整数规划简介

线性规划问题中有一部分问题要求有整数可行解和整数最优解，例如完成任务的人数、生产机器的台数、生产任务的分配、场址的选定等，都必须部分或者全部满足整数要求，这样的问题称为整数线性规划问题，简称整数规划问题。

1. 整数规划的引入

先举一些整数线性规划的实例。

【例 2.28】某商场决定：营业员每周连续工作 5 天后连续休息两天，轮流休息。根据统计，商场每天需要的营业员见表 2-26。

表 2-26　营业员需求表

星　期	一	二	三	四	五	六	日
需求人数	300	300	350	400	480	600	550

问商场人力资源部应如何安排每天的上班人数，使商场总的营业员最少。

解：设 $x_j\,(j=1,2,\cdots,7)$ 为休息两天后星期一到星期日开始上班的营业员，则星期一上班的人数包括星期一上班的营业员(连续工作 1 天)、星期四上班营业员(连续工作 5 天)、星期五上班的营业员(连续工作 4 天)、星期六上班的营业员(连续工作 3 天)和星期日上班的营业员(连续工作两天)，其他要求同理；该问题的目标函数为商场总的营业员最少，即要求 $x_1+x_2+x_3+x_4+x_5+x_6+x_7$ 为最小，所以该问题的线性规划模型为

$$\min z = x_1+x_2+x_3+x_4+x_5+x_6+x_7$$

$$\begin{cases} x_1+x_4+x_5+x_6+x_7 \geqslant 300 \\ x_1+x_2+x_5+x_6+x_7 \geqslant 300 \\ x_1+x_2+x_3+x_6+x_7 \geqslant 350 \\ x_1+x_2+x_3+x_4+x_7 \geqslant 400 \\ x_1+x_2+x_3+x_4+x_5 \geqslant 480 \\ x_2+x_3+x_4+x_5+x_6 \geqslant 600 \\ x_3+x_4+x_5+x_6+x_7 \geqslant 550 \\ x_j \geqslant 0 \text{且为整数}, j=1,2,\cdots,7 \end{cases}$$

由于所有变量都要求取整数，称它为纯整数规划问题。

【例 2.29】有时企业要作投资决策，就是对几个潜在的投资方案作出选择，例如投资决策可以是在可行的几个厂址中作出选择；或设备购置的选择；或对一组研究和发展项目作出决定。在这类决策问题中，问题是在“要”或者“不要”之间进行选择，如果令决策变量 x_j 是整数，且只取 0 或 1，分别表示不投资或者投资第 j 个方案，这种取值的变量为 0−1 变量。假定 c_j 代表第 j 项投资得到的收益，a_{ij} 是用于第 j 个方案的消耗第 i 项资源的数量，b_i 为第 i 种资源的限制，则上述问题所耗费资源约束为 $\sum\limits_{j=1}^{n} a_{ij}x_j \leqslant b_i$，企业投资是为了追求利润最大化，其目标函数为 $\sum\limits_{j=1}^{n} c_j x_j$。这样，该问题的规划模型为

北京大学出版社版权所有 禁止转载

$$\max z = \sum_{j=1}^{n} c_j x_j$$

$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j \leqslant b_i \ (i = 1, 2, \cdots, m) \\ x_j = 0 \text{ 或 } 1 \quad (j = 1, 2, \cdots, n) \end{cases}$$

由于所有的变量都只能取 0 或 1，这样的整数规划问题称为 0-1 规划。

【例 2.30】拟在某区域内建立某个商品的配送中心，如图 2.9 所示。有 3 个可供选择的配送点 $A_i (i = 1, 2, 3)$，容量为 a_i，不考虑固定资产投资；该区域有 4 个需求点 $B_j (j = 1, 2, 3, 4)$，各需求点的需求量为 d_j，配送点 A_i 到需求点 B_j 的运费为 c_{ij}，问：如何选择两个配送点，在满足需求的条件下，总费用最小。

解：令 $x_i (i = 1, 2, 3)$ 表示是否在 A_i 点设立配送中心，$x_i = 1$ 表示在 A_i 点设立配送中心，反之 $x_i = 0$；y_{ij} 表示从配送点 A_i 到需求点 B_j 的运量。

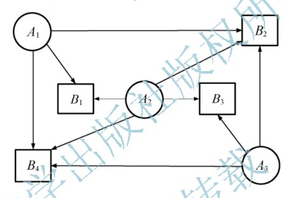


图 2.9　配送中心示意图

北京大学出版社版权所有 禁止转载

根据问题要求，三个可供选择的配送点选择两个的约束意义为 $x_1 + x_2 + x_3 = 2$；该区域的配送条件是要满足其需求，如需求点 B_1 的需求满足条件为 $y_{11} + y_{21} = d_1$，其他需求点同理；此外配送中心 A_1 还要受到其容量的限制，即 $y_{11} + y_{12} + y_{14} \leqslant a_1$，由于题中要求 3 个配送点要选择两个，或许 A_1 没有选到，那么 y_{11}、y_{12}、y_{14} 都应为零，将上面两种情况合并在一起，用一个约束可以表示为 $y_{11} + y_{12} + y_{14} \leqslant a_1 x_1$；目标函数为总费用最小，也就要求 $\sum_{i=1}^{3}\sum_{j=1}^{4} c_{ij} y_{ij}$ 最小，所以建立模型为

$$\min z = \sum_{i=1}^{3}\sum_{j=1}^{4} c_{ij} y_{ij}$$

$$\begin{cases} x_1 + x_2 + x_3 = 2 \\ y_{11} + y_{21} = d_1 \\ y_{12} + y_{22} + y_{32} = d_2 \\ y_{23} + y_{33} = d_3 \\ y_{14} + y_{24} + y_{34} = d_4 \\ y_{11} + y_{12} + y_{14} \leqslant a_1 x_1 \\ y_{21} + y_{22} + y_{23} + y_{24} \leqslant a_2 x_2 \\ y_{32} + y_{33} + y_{34} \leqslant a_3 x_3 \\ x_i \text{为} 0-1 \text{变量}, y_{ij} \geqslant 0 \end{cases}$$

在这个问题中，变量 $x_i(i=1,2,3)$ 只能取 0 或 1，变量 y_{ij} 为不小于 0 的数，这样的整数规划问题称为混合整数规划。

2. 整数规划的分类

一般整数规划分为两种类型，如果所有的变量都限制为非负整数，就称这类整数规划为纯整数规划，或整数规划，如例 2.28、例 2.29 是纯整数规划。纯整数规划的一个特殊情况是变量取值仅限于 0 或 1，该类问题称为 0−1 整数规划问题，也称为 0−1 规划，例 2.29 就是 0−1 规划。在整数规划中，如果一部分变量要求取整数而另一部分不一定要求取整数，称该类问题为混合整数规划问题，例 2.30 就是混合整数规划。

常用求解整数规划的方法有分枝定界法和割平面法，这两种方法书中不作介绍。对于特别的 0−1 规划问题的求解，可以采用匈牙利法，将在第 3 章讨论。如果在应用过程中遇到整数规划的问题，在要求不太严格的情况下，可以用其相应线性规划的最优解来近似，也可通过专用数学软件来求解。

2.4.2　线性整数规划的应用

这里介绍一些整数规划的实际应用，并建立其数学模型。

1. 固定费用问题

在讨论线性规划时，有些问题是要求使成本为最小，那时总设固定成本为常数，并在线性规划的模型中不必明显列出。但有些固定费用(固定成本)的问题(Fixed Cost Problem)不能用一般线性规划来描述，但可转换为混合整数规划来解决。

某工厂为了生产某种产品，有几种不同的生产方式可供选择，如选定的生产方式投资高(选购自动化程度高的设备)，由于产量大，因而分配到每件产品的变动成本就降低；反之，如选定的生产方式投资低，将来分配到每件产品的变动成本可能增加，所以必须全面考虑。

假设有 3 种方式可供选择，令 x_i 表示采用第 i 种方式时的产量；c_i 表示采用第 i 种方式时每件产品的变动成本；k_i 表示采用第 i 种方式时的固定成本。

(1) 为了说明成本的特点，暂不考虑其他约束条件。采用各种生产方式的总成本分别为

$$P_i=\begin{cases} k_i+c_ix_i, & \text{当}x_i>0 \\ 0, & \text{当}x_i=0 \end{cases} \quad i=1,2,3$$

(2) 在构建目标函数时，为了统一在一个问题中讨论，要引入 0−1 变量 y_j，令

$$y_j=\begin{cases} 1, & \text{当采用第 } j \text{ 种生产方式，} \quad \text{即}x_j>0\text{时} \\ 0, & \text{当不采用第 } j \text{ 种生产方式，} \quad \text{即}x_j=0\text{时} \end{cases}$$

于是成本的目标函数为 $\min z=(k_1y_1+c_1x_1)+(k_2y_2+c_2x_2)+(k_3y_3+c_3x_3)$

(3) 统一 x_j、y_j 的关系。当 $x_j>0$ 时，y_j 必须为 1；当 $x_j=0$ 时，y_j 必须为 0，为了统一这两种情况或状态，需引入一个充分大的数 M，表示为

$$x_j \leqslant y_jM, \quad j=1,2,3 \tag{2-29}$$

其中 M 是 x_j 取值的上界，可从题中条件得出。

北京大学出版社版权所有 禁止转载

【例 2.31】某农场有 100 公顷土地可用于发展生产，农场劳动力供给为 4 000 人·日。该农场种植三种作物：大豆、玉米和小麦，并饲养奶牛和鸡。种作物时不需要专门投资，而饲养动物建立牛栏需要花费 5 000 元，鸡舍需要花费 1 000 元。养奶牛时每头需拨出 1.5 公顷土地种饲草，并占用人工 70 人·日，年净收入 400 元/每头奶牛；养鸡时不占土地，需人工为每只鸡 0.5 人·日，年净收入为每只鸡 2 元。三种作物每年需要的人工及收入情况见表 2-27。

表 2-27　人工及收入情况表

	大　豆	玉　米	小　麦
劳动力/(人·日)	50	75	40
年净收入/(元/公顷)	175	300	120

试确定该农场的经营方案，使年净收入为最大。

解：如果设 x_1、x_2、x_3 分别表示种植大豆、玉米、小麦的面积，x_4、x_5 表示饲养奶牛、鸡的数量。考虑到饲养动物需要固定投资，引入两个 0-1 变量 y_1、y_2 表示是否饲养奶牛和鸡，即

$$y_1=\begin{cases}1, & \text{当饲养奶牛时}, \quad x_4>0\\ 0, & \text{当不饲养奶牛时}, \quad x_4=0\end{cases} \qquad y_2=\begin{cases}1, & \text{当饲养鸡时}, \quad x_5>0\\ 0, & \text{当不饲养鸡时}, \quad x_5=0\end{cases}$$

据题意：

(1) 占用土地的约束为　$x_1+x_2+x_3+1.5x_4 \leqslant 100$。

(2) 需要劳动力的要求为　$50x_1+75x_2+40x_3+70x_4+0.5x_5 \leqslant 4\,000$。

(3) 饲养或不饲养奶牛和鸡的约束为　$x_4 \leqslant y_1 M$；$x_5 \leqslant y_2 M$。

(4) 目标函数为扣除支出的总收益：

$$175x_1+300x_2+120x_3+400x_4+2x_5-5\,000y_1-1\,000y_2$$

于是该问题的线性规划模型可表示为

$$\max z=175x_1+300x_2+120x_3+400x_4+2x_5-5\,000y_1-1\,000y_2$$

$$\begin{cases} x_1 + x_2 + x_3 + 1.5x_4 \leqslant 100 \\ 50x_1 + 75x_2 + 40x_3 + 70x_4 + 0.5x_5 \leqslant 4\,000 \\ x_4 \leqslant y_1 M \\ x_5 \leqslant y_2 M \\ x_1, x_2, x_3, x_4, x_5 \geqslant 0;\ y_1, y_2\text{是0-1变量} \end{cases}$$

求解可得农场的经营方案为养牛 57 头，其他都不经营，年净收入为 17 800 元。

2. 多种类、多方案投资问题

【例 2.32】某校基金会有一笔数额为 5 000 万元的基金，打算将其存入银行。当前银行存款的利率期限结构见表 2-28。取款政策与银行的现行政策相同，定期存款不提前取，活期存款可任意支取。校基金会计划在 5 年内每年用部分本息奖励优秀师生，要求每年的奖金额为大致相同的整数，且在 5 年末仍保留原基金数额。校基金会希望获得最佳的基金使用计划，以提高每年的奖金额。帮助校基金会设计一个基金最佳使用方案，试建立其模型。

北京大学出版社版权所有 禁止转载

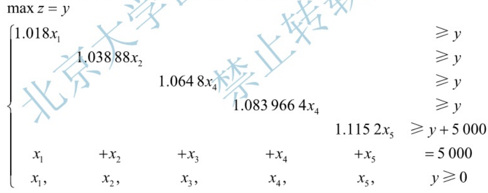
表 2-28　利率表

	活　期	半年期	一年期	二年期	三年期	五年期
银行存款税后年利率/(%)	0.792	1.664	1.800	1.944	2.160	2.304

解：参照存款年利率数据表 2-28 可知，定期存款年限越长，存款年利率越大。因此，在不影响奖金发放的情况下，应尽可能存年限较长的定期存款，这样才能获得较高的利息。所以此基金的最佳使用计划是：拿出一部分基金存入一年定期，一年后的本息全部用于发放第一年的奖金；再拿出一部分基金存入二年定期，二年后的本息全部用于发放第二年的奖金，以此类推，且每年发放奖金数额相同，最后一年得到的存入银行款项在发完奖金后仍然为基金总额。

假设每年发放奖学金一次，且均在年末发放；利率计算方式为单利；不考虑利息税。令每年发放的奖金额为 y，存期为 i 年的基金为 $x_i (i=1,2,3,4,5)$。根据题中要求，总基金为 5 000 万元的约束为 $x_1+x_2+x_3+x_4+x_5=5\,000$；第一年的奖金发放来源于存期为 1 年的存款及其利息，其约束为 $(1+0.018)x_1 \geqslant y$；以此类推，因为没有 4 年期的利息，这时存款方案有：先存 1 年再存 3 年、存两次 2 年、先存 3 年再存 1 年、先存 2 个 1 年再存 2 年、存 4 次一年等，通过各种方案的比较，可知存款方案为先存 3 年与再存 1 年的组合效益最高。第四年的奖金发放来源于先存 3 年与再存 1 年的存款及其利息，即 $(1+3\times 0.021\,6)(1+0.018)x_4 \geqslant y$；第五年奖金发放来源于存期为 5 年的存款及其利息，即 $(1+5\times 0.0230\,4)x_5 \geqslant y+5\,000$，目标函数为所发奖金为最大，即得该问题的数学模型为

$$\max z = y$$



$$\begin{cases} 1.018x_1 \geqslant y \\ 1.038\,88x_2 \geqslant y \\ 1.064\,8x_4 \geqslant y \\ 1.083\,966\,4x_4 \geqslant y \\ 1.115\,2x_5 \geqslant y+5\,000 \\ x_1+x_2+x_3+x_4+x_5 = 5\,000 \\ x_1,\ x_2,\ x_3,\ x_4,\ x_5,\ y \geqslant 0 \end{cases}$$

求解可得每年分配的奖金额 $y \approx 110$ 万元，每年的存款额可从约束条件中计算出来。

3. 运动员选拔问题

【例 2.33】篮球队需要选择 5 名队员组成出场阵容参加比赛。可候选的 8 名队员的身高和擅长位置见表 2-29。出场阵容应满足以下条件：

(1) 中锋只能有一个上场；

(2) 至少有一名后卫；

(3) 如 1 号或 4 号上场，则 6 号不上场；

(4) 2 号和 6 号至少保留一个不上场。

北京大学出版社版权所有 禁止转载

表 2-29　队员情况表

队　　员	身高/米	擅长位置	队　　员	身高/米	擅长位置
1	1.92	中锋	5	1.85	前锋
2	1.90	中锋	6	1.83	后卫
3	1.88	前锋	7	1.80	后卫
4	1.86	前锋	8	1.78	后卫

问应当选择哪 5 名队员上场，才能使出场队员平均身高最高？

解： 令 $x_i(i=1,2,\cdots,8)$ 表示是否选择第 i 号队员上场，$x_i=1$ 表示选择第 i 号队员上场，反之 $x_i=0$。

根据题知条件，得到：(1)的约束为 $x_1+x_2=1$；(2)的约束为 $x_6+x_7+x_8\geqslant 1$；(3)的约束为 $x_1+x_6\leqslant 1, x_4+x_6\leqslant 1$；(4)的约束 $x_2+x_6\leqslant 1$。建立规划模型为

$$\max z=\frac{1}{5}(1.92x_1+1.90x_2+1.88x_3+1.86x_4+1.85x_5+1.83x_6+1.80x_7+1.78x_8)$$

$$\begin{cases}\sum\limits_{i=1}^{8}x_i=5\\ x_1+x_2=1\\ x_6+x_7+x_8\geqslant 1\\ x_1+x_6\leqslant 1\\ x_4+x_6\leqslant 1\\ x_2+x_6\leqslant 1\\ x_1,\cdots,x_8\text{是0-1变量}\end{cases}$$

求解可得上场队员为 1、3、4、5、7，其平均身高为 1.862 米。

北京大学出版社版权所有
禁止转载

4. 背包问题

(二维背包问题)一个旅行者要在背包里装一些最有用的旅行物品。背包容积为 a，携带物品总质量最多为 b。现在有物品 n 种，第 i 种物品的体积为 a_i，质量为 b_i $(i=1,2,\cdots,m)$。为了比较物品的有用程度，假设第 i 种物品的有用程度为 c_i。若每种物品只能整件携带，每件物品都能放进背包中，并且不考虑物品放入背包后相互的间隙。问旅行者应携带哪几种物品才能使携带物品的总价值最大？

解： 令 x_i 为第 i 种物品的装入件数，则问题的数学模型是

$$\max z=\sum_{i=1}^{n}c_ix_i$$

$$\begin{cases}\sum\limits_{i=1}^{n}a_ix_i\leqslant a\\ \sum\limits_{i=1}^{n}b_ix_i\leqslant b\\ x_i\geqslant 0\text{且为整数},\ i=1,2,\cdots,n\end{cases}$$

【例 2.34】 有一艘货轮，它的容积为 5 400m^3，其最大允许载重量 3 000t，现有 3 种大批量的货物待运，已知有关数据见表 2-30。

表 2-30　商品明细表

商　　品	每件体积/(立方米/件)	每件重量/(吨/件)	运价/(元/件)
1	10	8	1 000
2	5	6	700
3	7	5	600

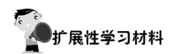
问该货轮应装载商品 1、2、3 各多少件，其运费收入为最大？

解： 设 x_i 为第 i 种物品的装入件数，$i=1,2,3$，根据题知条件，得容积约束为 $10x_1+5x_2+7x_3\leqslant 5\,400$；质量约束为 $8x_1+6x_2+5x_3\leqslant 3\,000$；运费收入为 $1\,000x_1+700x_2+600x_3$，那么其规划模型为

$$\max z=1\,000x_1+700x_2+600x_3$$

$$\begin{cases}10x_1+5x_2+7x_3\leqslant 5\,400\\ 8x_1+6x_2+5x_3\leqslant 3\,000\\ x_i\geqslant 0\text{且为整数，}\ i=1,2,3\end{cases}$$

求解，可得装载商品 1、2、3 分别为 375、0、0 件，其运费收入为 375 000 元。



扩展性学习材料

北京大学出版社版权所有 禁止转载

非线性规划

具有非线性约束条件或目标函数的数学规划，是运筹学的一个重要分支。非线性规划研究的是一个 n 元实函数在一组等式或不等式的约束条件下的极值问题，且目标函数和约束条件至少有一个是未知量的非线性函数。目标函数和约束条件都是线性函数的情形则属于线性规划。

非线性规划是 20 世纪 50 年代才开始形成的一门新兴学科。1951 年 H. W. 库恩和 A. W. 塔克发表的关于最优性条件(后来称为库恩－塔克条件)的论文是非线性规划正式诞生的一个重要标志。在 20 世纪 50 年代还得出了可分离规划和二次规划的 n 种解法，它们大都是以 G. B. 丹捷格提出的解线性规划的单纯形法为基础的。20 世纪 50 年代末到 60 年代末出现了许多解非线性规划问题的有效的算法，70 年代又得到进一步的发展。非线性规划在工程、管理、经济、科研、军事等方面都有广泛的应用，为最优设计提供了有力的工具。

本 章 小 结

本章学习了线性规划的最经典内容。首先从生产计划问题中引出了线性规划的一般模型及其标准型，通过对两个决策变量的问题示范了线性规划的图解方法，为了解线性规划各种解的情况做了直观准备。第二深入阐述了求解线性规划的单纯形方法，着重给出了单纯形法的理论基础和求解步骤，描述了实施其过程的单纯形表结构，并为寻找单纯形表的初始可行基给出了大 M 法。第三介绍了线性规划的对偶问题和灵敏度分析的有关内容，叙述了对偶理论和对偶单纯形法，为线性规划的求解指出了另一条途径。为使线性规划的理论应用更符合现实环境，刻画了线性规划各类参数的变化对其最优解的影响。第四介绍了整数规划的分类与其在实际问题中的具体应用，示范了一些整数规划模型的建立过程。



关键术语(中英文)

线性规划(Linear Programming)
目标函数(Objective Function)
价值向量(Price Coefficient Vector)
技术矩阵(Technique Matrix)
松弛变量(Slack Variable)
自由变量(Free Variable)
可行域(Feasible Area)
可行解(Feasible Solution)
基(Basis)
基础解(Basis Solution)
单纯形法(Simplex Method)
出基变量(Leaving Variable)
单纯形表(Simplex Tableau)
换基迭代(Basis Iteration)
人工变量(Artificial Variable)
弱对偶(Weak Duality)
影子价格(Shadow Price)
松弛互补(Complementary Slackness)
灵敏度分析(Sensitivity Analysis)
0-1 整数规划(Binary Integer Programming)
约束条件(Constraint)
决策变量(Decision Variable)
资源向量(Resource Vector)
标准型(Standard Form)
剩余变量(Surplus Variable)
图解法(Diagram Method)
解(Solution)
最优解(Optimal Solution)
基变量(Basic Variable)
基础可行解(Basic Feasible Solution)
顶点(Corner Point)
进基变量(Entering Variable)
检验数(Test Number)
大 M 法(Big M Method)
对偶问题(Dual Problem)
强对偶(Strong Duality)
差额成本(Reduced Cost)
对偶单纯形法(Duel Simplex Algorithm)
整数规划(Integer Programming)
混合整数规划(Mixed Integer Programming)

北京大学出版社版权所有 禁止转载



知识链接

线性规划之父——丹捷格

线性规划的诞生是在 1947 年的美国，33 岁的美国数学家 George Bernard Dantzig (丹捷格)提出了解决一种最优化问题的单纯形法，所解决的问题是美国空军军事规划时提出的，并给出了求解线性规划问题的单纯形算法。该方法奠定了线性规划的基础，使得经济学、环境科学、统计学应用等学科获得了迅速发展。Dantzig 也因而被誉为“线性规划之父”。

这位斯坦福大学的数学专家，在他的一生中得到过无数的数学奖项，而且被公认为现代运筹学的创始人之一。Dantzig 在运筹学建树极高，获得了包括“冯·诺伊曼理论奖”在内的诸多奖项。他在 *Linear Programming and Extensions* 中研究了线性编程模型，为计算机语言的发展作出了不可磨灭的贡献。由于他的研究，领导美国的运筹学在这方面处在相对领先的位置；而且线性规划迅速地得到工商业界的应用，并且为第二次世界大战以后的经济重建作出了伟大的贡献。

Dantzig 的父亲是大学数学教授，曾在法国师从著名的科学家 Henri Poincare。但是 Dantzig 直到上初中时，对数学仍不感兴趣，甚至在初中三年级时，代数成

绩还不及格。对于这样的结果，Dantzig 非常内疚，他感到愧对自己的数学家父亲，于是发奋努力，很快就发现其实数学并不难，逐渐地建立了自信。上高中时 Dantzig 对父亲的数学题库非常着迷，他解出了所有的题目。Dantzig 曾经这样回忆自己的父亲："在我还是个中学生时，他就让我做几千道几何题……解决这些问题的大脑训练是父亲给我的最好礼物。这些几何题，在发展我分析能力的过程中，起了最最重要的作用。"



习题 2

2.1 填空题

(1) 在线性规划问题中，称满足所有约束条件方程和非负限制的解为__________。

(2) 在求解线性规划的算法中，当__________时采用单纯形法；当__________时采用对偶单纯形法。(**提示：**考虑 $\boldsymbol{B}^{-1}\boldsymbol{b}$，$\boldsymbol{C}-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}$ 的符号)

(3) 在线性规划问题中，基变量所对应的检验数一定为__________。

(4) 线性规划检验数 σ_j 的经济含义是__________。

(5) 在资源优化的线性规划问题中，某资源有剩余，则该资源影子价格等于__________。

2.2 判断题(正确的打 √，错误的打×)

(1) 影子价格就是资源的价格。 (　　)

(2) 线性规划问题的每一个基解对应可行域的一个顶点。 (　　)

(3) 图解法与单纯形法，虽然求解的形式不同，但从几何上理解，两者是一致的。(　　)

(4) 线性规划的最优解一定是基本最优解。 (　　)

(5) 互为对偶问题，或者同时都有最优解，或者同时都无最优解。 (　　)

2.3 将下列线性规划问题变换成标准型

(1) $\max z = 2x_1 + x_2 + 3x_3 + x_4$

$$\begin{cases} x_1 + x_2 + x_3 + x_4 \leqslant 7 \\ 2x_1 - 3x_2 + 5x_3 = -8 \\ x_1 \quad\quad - 2x_3 + 2x_4 \geqslant 1 \\ x_1, x_3 \geqslant 0, x_2 \leqslant 0, x_4\text{无约束} \end{cases}$$

(2) $\min z = 2x_1 - x_2 + 2x_3$

$$\begin{cases} -x_1 + x_2 + x_3 = 4 \\ -x_1 + x_2 - x_3 \leqslant 6 \\ x_1 \leqslant 0, x_2 \geqslant 0, x_3\text{无约束} \end{cases}$$

2.4 考虑线性规划问题

$$\max z = c_1 x_1 + x_2$$
$$\begin{cases} x_1 + x_2 \leqslant 6 \\ x_1 + 2x_2 \leqslant 10 \\ x_1 \geqslant 0,\ x_2 \geqslant 0 \end{cases}$$

其中 $c_1 \in (-\infty, +\infty)$ 为参数，试用图解法确定其最优解的情况。

2.5 对于以下线性规划问题(如图 2.10 所示)

$$\max z = -x_1 + 2x_2$$
$$\begin{cases} 2x_1 + 3x_2 \leqslant 12 \\ 3x_1 + x_2 \leqslant 6 \\ -x_1 + 3x_2 \leqslant 3 \\ x_1,\ x_2 \geqslant 0 \end{cases}$$

北京大学出版社版权所有 禁止转载



3 个约束对应的松弛变量分别为 x_3, x_4, x_5；3 个约束条件对应的对偶变量分别为 w_3, w_4, w_5。试回答以下问题。

(1) 问题的可行域。

(2) 问题的基本解。

(3) 问题的基本可行解。

(4) 问题的最优解。

(5) G 点对应的解中，大于 0 的变量为哪些？等于 0 的变量为哪些？小于 0 的变量为哪些？

(6) F 点对应的基变量为哪些？非基变量为哪些？

(7) E 点对应的基变量为哪些？非基变量为哪些？

(8) 从 F 到 E 的单纯形迭代，进基变量为哪些？离基变量为哪些？

(9) E 点对应的对偶变量，大于 0 的为哪些？等于 0 的为哪些？小于 0 的为哪些？

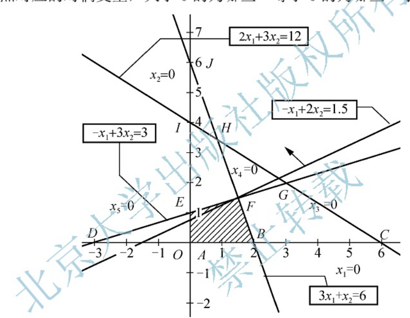


图 2.10　题 2.5 图

北京大学出版社版权所有 禁止转载

2.6 求解线性规划问题时，使用单纯形法求解，已知初始单纯形表和终表，见表 2-31，试求解以下问题。

(1) 在初始单纯形表中找出最优基 $\boldsymbol{B}$，在单纯形终表里找出 $\boldsymbol{B}^{-1}$；

(2) 完成单纯形终表；

(3) 给出最优解与最优值。

表 2-31　初始单纯形表和终表

	$\boldsymbol{C}_B$	$\boldsymbol{X}_B$	$\boldsymbol{B}^{-1}\boldsymbol{b}$	2 x_1	−1 x_2	1 x_3	0 x_4	0 x_5	0 x_6
	0	x_4	60	3	1	1	1	0	0
初	0	x_5	10	1	−1	2	0	1	0
表	0	x_6	20	1	1	−1	0	0	1
		σ_j		2	−1	1	0	0	0

续表

	C_B	X_B	$B^{-1}b$	2	−1	1	0	0	0
				x_1	x_2	x_3	x_4	x_5	x_6
终表		x_4						−1	−2
		x_1						1/2	1/2
		x_2						−1/2	1/2
		σ_j							

2.7 设有线性规划问题

$$\max z = 5x_1 + 2x_2 + 3x_3$$

$$\begin{cases} x_1 + 5x_2 + 2x_3 \leqslant b_1 \\ x_1 - 5x_2 - 6x_3 \leqslant b_2 \\ x_1, \ x_2, \ x_3 \geqslant 0 \end{cases}$$

其中b_1、b_2是常数，且已知该线性规划的最终单纯形表，见表2-32。

表2-32　最终单纯形表

X_B	$B^{-1}b$	5	2	3	0	0
		x_1	x_2	x_3	x_4	x_5
x_1	30	1	a	2	1	0
x_5	10	0	c	−8	−1	1
检验数σ		0	d	−7	e	f

试求：(1) 确定b_1, b_2；

(2) 确定a, c, d, e, f的值；

(3) 求相应的对偶规划的最优解。

2.8 表2-33给出某最大化线性规划问题的单纯形表。试确定表中参数a, d, e, f, g为何取值范围时，满足以下要求：

(1) 表中解为唯一最优解；

(2) 表中解为多重最优解之一；

(3) 该问题目标函数值无界；

(4) 以x_4替代x_2后，目标函数值将会增加。

表2-33　单纯形表

X_B	$B^{-1}b$	x_1	x_2	x_3	x_4	x_5	x_6
x_2	a	0	1	g	2	0	3
x_1	1	1	0	d	1	0	4
x_5	6	0	0	−2	−4	1	1
检验数σ		0	0	e	f	0	−2

北京大学出版社版权所有 禁止转载

2.9 应用单纯形法求解下列线性规划问题，并指出问题的解属于哪一类。

(1) $\max z = 3x_1 + 5x_2$

$$\begin{cases} x_1 \leqslant 4 \\ 2x_2 \leqslant 12 \\ 3x_1 + 2x_2 \leqslant 18 \\ x_1, x_2 \geqslant 0 \end{cases}$$

(2) $\max z = 2x_1 - x_2 + x_3$

$$\begin{cases} 3x_1 + x_2 + x_3 \leqslant 60 \\ x_1 - x_2 + 2x_3 \leqslant 10 \\ x_1 + x_2 - x_3 \leqslant 20 \\ x_1, x_2, x_3 \geqslant 0 \end{cases}$$

2.10 利用大 M 法或对偶单纯形法求解以下线性规划。

(1) $\min w = 5x_1 + 2x_2 + 4x_3$

$$\begin{cases} 3x_1 + x_2 + 2x_3 \geqslant 4 \\ 6x_1 + 3x_2 + 5x_3 \geqslant 10 \\ x_1, x_2, x_3 \geqslant 0 \end{cases}$$

(2) $\min z = 4x_1 + 12x_2 + 18x_3$

$$\begin{cases} x_1 + 3x_3 \geqslant 3 \\ 2x_2 + 2x_3 \geqslant 5 \\ x_1, x_2, x_3 \geqslant 0 \end{cases}$$

2.11 对于问题

$$\max z = \boldsymbol{CX}$$

$$\begin{cases} \boldsymbol{AX} = \boldsymbol{b} \\ \boldsymbol{X} \geqslant \boldsymbol{0} \end{cases}$$

(1) 设其最优解为 $\boldsymbol{X}^*$，当 $\boldsymbol{C}$ 改为 $\overline{\boldsymbol{C}}$ 时，最优解为 $\overline{\boldsymbol{X}}$，则 $(\boldsymbol{C} - \overline{\boldsymbol{C}})(\boldsymbol{X}^* - \overline{\boldsymbol{X}}) \geqslant 0$。

(2) 如果 $\boldsymbol{X}_1, \boldsymbol{X}_2$ 均为最优解，则对于 $\alpha \in [0,1]$，$\alpha\boldsymbol{X}_1 + (1-\alpha)\boldsymbol{X}_2$ 均为最优解。

2.12 给出线性规划问题

$$\max z = x_1 + 2x_2 + x_3$$

$$\begin{cases} x_1 + x_2 - x_3 \leqslant 2 \\ x_1 - x_2 + x_3 = 1 \\ 2x_1 + x_2 + x_3 \geqslant 2 \\ x_1 \geqslant 0,\ x_2 \leqslant 0 \end{cases}$$

试求：(1)写出其对偶问题；(2)利用对偶问题证明原问题目标函数值 $z \leqslant 1$。

2.13 考虑下列线性规划

$$\max z = 2x_1 + x_2 + 3x_3$$

$$\begin{cases} x_1 + x_2 + 2x_3 \leqslant 5 \\ 2x_1 + 3x_2 + 4x_3 = 12 \\ x_1, x_2, x_3 \geqslant 0 \end{cases}$$

(1) 写出其对偶问题；

(2) 已知(3，2，0)是上述原问题的最优解，根据互补松弛定理，求出对偶问题的最优解；

(3) 如果上述规划中的第一个约束为资源约束，写出这种资源的影子价格。

2.14 考虑下面线性规划问题

$$\min w = 4x_1 + 6x_2 + 18x_3$$

$$\begin{cases} x_1 + 3x_3 \geqslant 3 \\ x_2 + 2x_3 \geqslant 5 \\ x_1, x_2, x_3 \geqslant 0 \end{cases}$$

北京大学出版社版权所有 禁止转载

试回答以下问题：

(1) 用对偶单纯形法求解；

(2) 求其对偶问题，并用单纯形法求解；

(3) 比较原规划与其对偶问题求解的终表，检验数和另一个最优解的关系。

2.15 已知线性规划问题

$$\max z = 10x_1 + 5x_2$$
$$\begin{cases} 3x_1 + 4x_2 \leqslant 9 \\ 5x_1 + 2x_2 \leqslant 8 \\ x_1,\ x_2 \geqslant 0 \end{cases}$$

用单纯形法求得最终单纯形表见表 2-34，试用灵敏度分析的方法判断：

(1) 目标函数系数 c_1 或 c_2 分别在什么范围内变动，现最优解不变？

(2) 约束条件右端常数 b_1、b_2 中，当保持一个不变时，另一个在什么范围内变化，现有的最优基不变？

(3) 问题的目标函数变为 $\max z = 12x_1 + 4x_2$ 时，最优解有什么变化？

(4) 约束条件右端常数由 $\begin{bmatrix} 9 \\ 8 \end{bmatrix}$ 变为 $\begin{bmatrix} 11 \\ 19 \end{bmatrix}$ 时，最优解有什么变化？

表 2-34 最终单纯形表

C_B	X_B	$B^{-1}b$	2	-1	1	0
			x_1	x_2	x_3	x_4
5	x_2	3/2	0	1	5/14	−3/14
10	x_1	1	1	0	−1/7	2/7
σ_j			0	0	−5/14	−25/14

北京大学出版社版权所有 禁止转载

2.16 某厂生产 A、B、C 三种产品，其所需资源及产品利润有关数据见表 2-35。要求：

(1) 确定获利最大的产品生产计划；

(2) 产品 A 的利润在什么范围内变动时，上述最优计划不变；

(3) 如果设计一种新产品 D，单位劳动力消耗为 8 单位，材料消耗为 2 单位，每件可获利 3 元，问该种产品是否值得生产？

(4) 如果劳动力数量不增，材料不足时可以从市场购买，每单位 0.4 元。问该厂是否需要购进原材料从而扩大生产，其购买量多少为宜？

表 2-35 某厂生产 A、B、C 三种产品所需资源及产品利润

产品 / 消耗定额 / 资源	A	B	C	可用单位
劳动力	6	3	5	45
材料	3	4	5	30
产品利润/(元/件)	3	1	4	

2.17 某钻井队要从以下 10 个可供选择的井位中确定 5 个钻井探油，使总的钻探费用为最小。若 10 个井位的代号为 $s_1,s_2,\cdots,s_{10}$，相应的钻探费用为 $c_1,c_2,\cdots,c_{10}$，并且井位选择上要满足以下限制条件：

(1) 或选择 s_1 和 s_7，或选择钻探 s_9；

(2) 选择了 s_3 或 s_4 就不能选 s_5，反过来也一样；

(3) 在 s_5,s_6,s_7,s_8 中最多只能选两个；试建立这个问题的整数规划模型。

2.18 某科学实验卫星拟从下列仪器装置中选若干件装上，有关数据见表 2-36。要求：

(1) 装入卫星的仪器装置的总体积不超过 V，总质量不超过 W；

(2) A_1 与 A_3 中至少安装一件；

(3) A_2 与 A_4 中至少安装一件；

(4) A_5 与 A_6 或者都安上，或者都不安。

安装的总的目标是使装上去的仪器装置使该科学卫星发挥最大的实验价值。试建立这个问题的数学模型。

表 2-36　各仪器有关数据

仪器代号	体　积	质　量	价　值
A_1	v_1	w_1	c_1
A_2	v_2	w_2	c_2
A_3	v_3	w_3	c_3
A_4	v_4	w_4	c_4
A_5	v_5	w_5	c_5
A_6	v_6	w_6	c_6

北京大学出版社版权所有
禁止转载



实际操作训练

某大型超市连锁公司欲在某市的东、南、西、北 4 区建立新的超市，拟议中有 10 个位置 $A_i(i=1,2,\cdots,10)$ 可供选择。考虑到各地区居民的消费水平及居民居住密集度，规定：在东区由 A_1,A_2,A_3 三个点至多选择两个，在西区由 A_4,A_5 两个点至少选择一个；在南区由 A_6,A_7 两个点至少选择一个；在北区由 A_8,A_9,A_{10} 三个点至多选择两个。A_i 各点的投资额及每年可获利润由于地点不同都是不一样的，预测情况见表 2-37。现有资金 720 万元，问应选择哪几个销售点，可使年利润为最大？

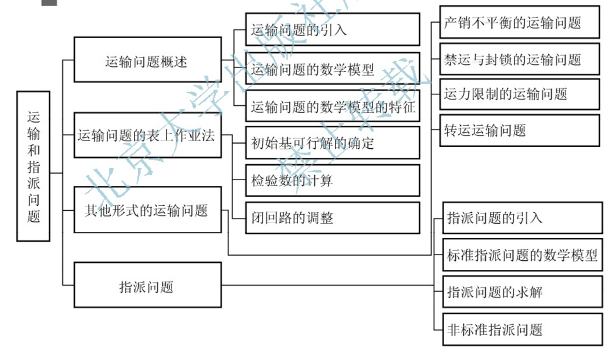
表 2-37　投资利润表

位　置	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
投资额 v_i	100	120	150	80	70	90	80	140	160	180
利润 c_i	36	40	50	22	20	30	25	48	58	61

第 3 章　运输和指派问题

北京大学出版社版权所有
禁止下载

【本章知识架构】



【本章教学目标与要求】

- 掌握产销平衡运输问题的数学模型及其特点。
- 掌握运输问题的表上作业法，包括初始调运方案的确定、检验数的计算、运输方案的调整方法。
- 掌握产销不平衡运输问题转化为产销平衡问题的处理办法；掌握运输问题在实践中的典型应用。
- 掌握标准指派问题的求解方法，会将各种非标准指派问题转化为标准指派问题。



运储物流的运输问题

运输成本占物流总成本的 35%～50%，占商品价格的 4%～10%，运输对物流总成本的节约具有举足轻重的作用。运储物流在物流运输管理中要着重考虑运输方式的选择、运输路线的选择、编制运输计划等问题。

运输方式合适与否决定了运输时间的长短与运输成本的高低，各种运输工具都有其使用的优势领域，对运输工具进行优化选择，按运输工具特点进行装卸运输作业，最大限度地发挥所用运输工具的作用；选择运输路线要与交通运输工具结合起来，尽量安排直达运输，以减少运输装卸、转运环节，缩短运输时间；编制运输计划还要从全局出发，深入调查研究，综合平衡，积极组织计划运输、合理运输、直达运输、均衡运输，按照成本最低的原则来制订合理的计划。

资料来源：http://www.baoyuntong.com.

线性规划在物流活动有着广泛的应用，本章介绍两类特别重要的线性规划问题：一是运输问题；二是指派问题，它们具有相似的模型结构，其内容丰富，求解方法独特。

3.1　运输问题概述

运输问题的典型提法是将某种物质从若干个产地调运到若干个销地，已知每个产地的产量和每个销地的销量，如何在许多可行调运方案中选择一个总运费最少的调运方案。根据总产量与总销量是否相等的数量关系，运输问题通常可划分为产销平衡(相等)和产销不平衡(不相等)两大类别。产销平衡的运输问题主要在本节介绍，产销不平衡的运输问题将在后面节中讨论。

3.1.1　运输问题的引入

在生产、交换活动中，不可避免地要进行物资调运工作。某时期内将生产基地的煤、钢铁、粮食、矿砂、木材等各类物资，分别运送到需要这些物资的地区。

【例 3.1】 某物流公司从两个产地 A_1(内蒙古)、A_2(山西)将煤炭运往 3 个销地 B_1(北京)、B_2(山东)、B_3(上海)，各产地的产量、各销地的销量、各产地运往各销地的每单位煤炭运费数据见表 3-1，问：应如何调运煤炭可使总运输费用最小？

表 3-1　单位运价表

销　地 / 产　地	B_1	B_2	B_3	产　量
A_1	6	4	6	200
A_2	6	5	5	300
销　量	150	150	200	500

注：上面都是虚拟数据。

北京大学出版社版权所有 禁止转载

解：此为产销平衡的运输问题(总产量=总销量)。

每个调运方案都要从产地调运若干煤炭到达销地，并满足各销地的需求，所以调运方案都落实在实际的调运数量上，即决策变量就是从各个产地调运多少煤炭到达销地，为表明这些调运数量的关系，特设 x_{ij} 为从产地 A_i $(i=1,2)$ 运往销地 B_j $(j=1,2,3)$ 的运输量，得到表 3-2。

表 3-2　运输问题变量

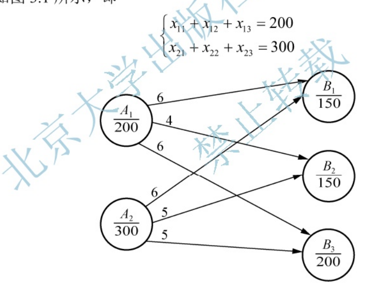
销　地 产　地	B_1	B_2	B_3	产　量
A_1	x_{11}	x_{12}	x_{13}	200
A_2	x_{21}	x_{22}	x_{23}	300
销　量	150	150	200	500

从而，目标函数表达式为：$6x_{11}+4x_{12}+6x_{13}+6x_{21}+5x_{22}+5x_{23}$。

1. 约束条件

(1) 产地约束。每个产地的物资全部运出，每个产地运出的物资量正好等于该产地的可供应量，如图 3.1 所示，即

$$\begin{cases} x_{11}+x_{12}+x_{13}=200 \\ x_{21}+x_{22}+x_{23}=300 \end{cases}$$



北京大学出版社版权所有　禁止转载

图 3.1　运输问题网络图

(2) 销地约束。每个销地的需求全部满足，运入每个收点的物资量正好等于该销地的需求量，即

$$\begin{cases} x_{11}+x_{21}=150 \\ x_{12}+x_{22}=150 \\ x_{13}+x_{23}=200 \end{cases}$$

(3) 蕴含约束。运输的煤炭量不能是负数，即决策变量的取值是非负的，即有 $x_{ij}\geqslant 0\,(i=1,2;j=1,2,3)$。

综上所述，该问题的数学模型为

$$\min z = 6x_{11} + 4x_{12} + 6x_{13} + 6x_{21} + 5x_{22} + 5x_{23}$$

$$\begin{cases} x_{11} + x_{12} + x_{13} = 200 \\ x_{21} + x_{22} + x_{23} = 300 \\ x_{11} + x_{21} = 150 \\ x_{12} + x_{22} = 150 \\ x_{13} + x_{23} = 200 \\ x_{ij} \geqslant 0 \quad (i = 1, 2;\ j = 1, 2, 3) \end{cases}$$

该问题显然是一个线性规划模型，其系数矩阵 $\boldsymbol{A}$ 见表 3-3。

表 3-3　运输问题的系数矩阵

	x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	x_{23}	
	1	1	1				产地约束
				1	1	1	
$\boldsymbol{A}$=	1			1			销地约束
		1			1		
			1			1	

2. 运输问题系数矩阵的特征

(1) 共有 $m+n$ 行，分别表示各产地和销地；$m \times n$ 列，分别表示各决策变量的系数列。

(2) 每列只有两个 1，其余为 0，表示只有一个产地和一个销地被使用，由 x_{ij} 的行列下标所决定，这是运输问题表上作业法的由来。

北京大学出版社版权所有 禁止转载

3.1.2　运输问题的数学模型

一般地，产销平衡的运输问题可以表述为：设有 m 个地点(称为产地或发地) $A_1, A_2, \cdots, A_m$ 的某种物资调至 n 个地点(称为销地或收地) $B_1, B_2, \cdots, B_n$，各个产地需要调出的物资量分别为 $a_1, a_2, \cdots, a_m$ 单位，各个销地需要调进的物资量分别为 $b_1, b_2, \cdots, b_n$ 单位，且各个发点的供应量之和等于各个收点的需求量之和。已知每个发点 A_i 到每个收点 B_j 的物资单位调运价格为 c_{ij}。现问如何安排调运，才能使总运费最小。

运输问题的上述相关信息可以放在表 3-4，称之为单位运价表。

表 3-4　单位运价表

销　地 产　地	B_1	B_2	…	B_n	供应量
A_1	c_{11}	c_{12}	…	c_{1n}	a_1
A_2	c_{21}	c_{22}	…	c_{2n}	a_2
⋮	⋮	⋮		⋮	⋮
A_m	c_{m1}	c_{m2}	…	c_{mn}	a_m
需求量	b_1	b_2	…	b_n	$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$

同上，若设 x_{ij} 表示从产地 A_i 运往 B_j 销地的运量($i=1,2,\cdots,m;\ j=1,2,\cdots,n$)，就得运输问题的决策变量列表，见表 3-5；当 x_{ij} 取具体数值时，该表也称为产销平衡表。

表 3-5 产销平衡表

销地 产地	B_1	B_2	…	B_n	供应量
A_1	x_{11}	x_{12}	…	x_{1n}	a_1
A_2	x_{21}	x_{22}	…	x_{2n}	a_2
⋮	⋮	⋮		⋮	⋮
A_m	x_{m1}	x_{m2}	…	x_{mn}	a_m
需求量	b_1	b_2	…	b_n	$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$

于是产销平衡运输问题的数学模型为

$$\min z = \sum_{j=1}^{n}\sum_{i=1}^{m} c_{ij}x_{ij}$$

$$\begin{cases} \sum_{j=1}^{n} x_{ij} = a_i,\ i=1,2,\cdots,m \\ \sum_{i=1}^{m} x_{ij} = b_j,\ j=1,2,\cdots,n \\ x_{ij} \geqslant 0,\ i=1,2,\cdots,m;\ j=1,2,\cdots,n \end{cases} \tag{3-1}$$

由式(3-1)可知，在运输问题的数学模型中，有 $m\times n$ 个决策变量，$m+n$ 个约束方程。若记

$$\boldsymbol{X} = [x_{11}, x_{12}, \cdots, x_{1n}, x_{21}, x_{22}, \cdots, x_{2n}, \cdots, x_{m1}, x_{m2}, \cdots, x_{mn}]^{\mathrm{T}}$$
$$\boldsymbol{C} = [c_{11}, c_{12}, \cdots, c_{1n}, c_{21}, c_{22}, \cdots, c_{2n}, \cdots, c_{m1}, c_{m2}, \cdots, c_{mn}]$$
$$\boldsymbol{A} = [P_{11}, P_{12}, \cdots, P_{1n}, P_{21}, P_{22}, \cdots, P_{2n}, \cdots, P_{m1}, P_{m2}, \cdots, P_{mn}]$$
$$\boldsymbol{b} = [a_1, a_2, \cdots, a_m, b_1, b_2, \cdots, b_n]^{\mathrm{T}}$$

则平衡运输问题模型写成矩阵形式为

$$\min z = \boldsymbol{CX}$$
$$\begin{cases} \boldsymbol{AX} = \boldsymbol{b} \\ \boldsymbol{X} \geqslant \boldsymbol{0} \end{cases} \tag{3-2}$$

在实际问题建立运输问题模型时，还会出现以下一些变化。

(1) 有时问题表面看不是运输问题，但其仍要求费用最低或要求目标函数(利润或营业额)最大化，仍可看成运输问题。

(2) 当某些运输线路上的能力有限制时，模型中可直接加入(等式或不等式)约束。

3.1.3 运输问题的数学模型的特征

运输问题是一类线性规划问题，其目标函数一般为求总运费的最小值。根据线性规划的有关理论，如果它的最优解存在，一定可以在基可行解中找到，因而需先考察运输问题式(3-1)约束方程组系数矩阵的秩 $r(\boldsymbol{A})$。

北京大学出版社版权所有 禁止转载

定理 3.1 运输问题模型(3-1)系数矩阵的秩 $r(\boldsymbol{A}) = m+n-1$。

考虑式(3-1)中的 $m+n$ 个约束条件，将前 m 个约束条件相加，得

$$\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} = \sum_{i=1}^{m} a_i$$

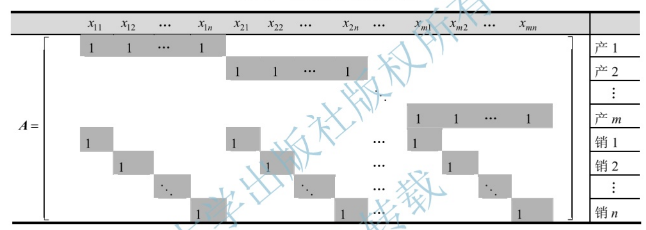
将后 n 个约束条件相加，得

$$\sum_{j=1}^{n}\sum_{i=1}^{m} x_{ij} = \sum_{j=1}^{n} b_j$$

由于产销平衡问题中，有 $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$，所以这 $m+n$ 个约束不是独立的。

接着考虑式(3-1)的系数矩阵，见表 3-6。

表 3-6 运输问题的系数矩阵



	x_{11}	x_{12}	⋯	x_{1n}	x_{21}	x_{22}	⋯	x_{2n}	⋯	x_{m1}	x_{m2}	⋯	x_{mn}	
	1	1	⋯	1										产 1
					1	1	⋯	1						产 2
									⋱					⋮
$\boldsymbol{A}=$										1	1	⋯	1	产 m
	1				1				⋯	1				销 1
		1				1			⋯		1			销 2
			⋱				⋱		⋯			⋱		⋮
				1				1	⋯				1	销 n

由此可见在运输问题系数矩阵中，决策变量 x_{ij} 对应的系数列向量可表示为

$$P_{ij} = (0 \quad \cdots \quad 1 \quad \cdots \quad 1 \quad \cdots \quad 0)^{\mathrm{T}} = \boldsymbol{e}_i + \boldsymbol{e}_{m+j}$$

其中：$\boldsymbol{e}_i$ 表示 $m+n$ 维单位列向量，它的第 i 个元素为 1。

注意到在 $\boldsymbol{A}$ 中去掉第 $m+n$ 行而取出第 1 行，第 2 行，⋯，第 $m+n-1$ 行，又取出与 $x_{1n}, x_{2n}, \cdots, x_{mn}, x_{11}, x_{12}, \cdots, x_{1,n-1}$ 所对应的列，则由这些取出行和列的交叉处的元素构成系数矩阵 $\boldsymbol{A}$ 的一个 $m+n-1$ 阶子式，因此 $r(\boldsymbol{A}) = m+n-1$，即运输问题基变量的个数为 $m+n-1$ 个。

运输问题是特殊的线性规划，单纯形法仍适合于运输问题。为此还要了解运输问题基可行解的性质，为说明其基本可行解的特征，引入闭回路的概念，下面的讨论都建立在表 3-5 的基础上。

定义 3.1 闭回路：在表 3-5 的决策变量表中，凡是能够排列成下列形式的

$$x_{ab},\ x_{ad},\ x_{cd},\ x_{ce},\ \cdots,\ x_{st},\ x_{sb} \tag{3-3}$$

或

$$x_{ab},\ x_{cb},\ x_{cd},\ x_{ed},\ \cdots,\ x_{st},\ x_{at} \tag{3-4}$$

式中：$a, d, \cdots, s$ 各不相同；$b, c, \cdots, t$ 各不相同，称这些变量的集合为一闭回路，并将式(3-3)和式(3-4)中的决策变量称为该闭回路的顶点。

例如 x_{13}，x_{16}，x_{36}，x_{34}，x_{24}，x_{23}；x_{23}，x_{53}，x_{55}，x_{45}，x_{41}，x_{21} 等都是闭回路，这些决策变量就是闭回路的顶点。

北京大学出版社版权所有 禁止转载

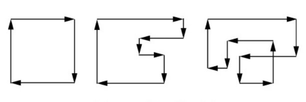


图 3.2　闭回路示例

若把闭回路的各变量格看做顶点，在表中可以画出如图 3.2 所示的闭回路，根据定义 3.1 可以看出闭回路的一些明显特点，闭回路是一个具有以下条件顶点格子的集合。

(1) 每一个顶点格子都是转角点。

(2) 每一行(或列)若有闭回路的顶点，则必有两个顶点。

(3) 每两个顶点格子的连线都是水平的或垂直的。

(4) 闭回路中顶点的个数必为偶数。

决策变量的闭回路具有定理 3.2 给出的性质。

定理 3.2 对于产销平衡运输问题的闭回路来说，具有以下结论。

(1) 该问题 $m+n-1$ 个变量构成基变量的充要条件是这些变量不包含任何闭回路。

(2) 给定一组基变量，那么从表 3-5 中可找出唯一一个从任意非基变量出发，经过基变量为顶点，又回到该非基变量的闭回路。

事实上，闭回路是一个简化的局部调运方案，反映了全局调运方案是否最优。定理 3.2 给出了运输问题基本解的重要性质，为寻求运输问题的基本可行解提供了依据。与一般的线性规划问题有所不同，产销平衡的运输问题总是存在可行解，且目标函数值有界，故运输问题必有最优解。

北京大学出版社版权所有 禁止转载

3.2　运输问题的表上作业法

运输问题作为一类特殊的线性规划问题，在求解时仍可采用单纯形法的计算步骤，但因为运输决策变量有两个下标，可以在单位运价表和产销平衡表中进行基变量与非基变量的换基迭代，再加上运输问题模型系数矩阵的特征，早期的研究者提出了专门针对运输问题的单纯形法——表上作业法。

表上作业法主要包括以下几个步骤。

第 1 步，求一个初始基可行解，即确定初始调运方案。

第 2 步，计算检验数并判断是否得到最优解。常见的求检验数的方法有闭回路法和位势法，求出检验数后可根据最优性条件来判定这个基可行解是不是最优的？若是则迭代停止；否则转下一步。

第 3 步，换基迭代，即调整运量。选一个变量出基，对原运输方案进行调整得到新的调运方案，改进当前方案，返回第 2 步，直至求出最优解为止。

3.2.1　初始基可行解的确定

确定初始可行解常用的方法有西北角法、最小元素法和差值法(Vogel 法)。

1. 西北角法

从产销平衡表的西北角(左上角)第一格开始，按集中供应的原则，依次安排调运量，即分析对应产地和销地的供需数量关系，尽最大可能满足需求，若某行(列)的产量(销量)已满足，则把该行(列)的其他格划去；接着从未被划线的运价中再找出西北角的方格，重复上述操作，直至得到一个基本可行解。

【例 3.2】设某种产品有 A_1,A_2,A_3 共 3 个生产厂商，联合供应 B_1,B_2,B_3,B_4 共 4 个销地，其供应量、需要量和单位产品的运输成本见表 3-7，试求一调运方案。

表 3-7　运输成本表

销地 / 产地	B_1	B_2	B_3	B_4	供　应　量
A_1	3	11	3	10	7
A_2	1	9	2	8	4
A_3	7	4	10	5	9
需求量	3	6	5	6	20

解：考虑表 3-7 的西北角——左上角格，其对应的决策变量 x_{11}，其调运量需从供应地 A_1 调运，满足销地 B_1 的需求，比较其供应与需求，尽可能最大限度地满足销地的需求，即 $x_{11}=\min\{3,7\}$，将数值 3 填入该方格(见表 3-8)。这样 A_1 发点供应量剩余 7−3=4，B_1 收点的需求量已经满足，因此 x_{21}、x_{31} 就不需调运物质，即第一列其他各方格变量都取零值，从而第一列在后面分析中不再考虑，划去第一列(见表 3-9)。

表 3-8　西北角法一

销地 / 产地	B_1	B_2	B_3	B_4	供　应　量
A_1	3 3	11	3	10	7
A_2	1 0	9	2	8	4
A_3	7 0	4	10	5	9
需求量	3	6	5	6	20

为简化表格形式，当决策变量取值为零时，其值在产销平衡表中省略，以后将取值为零的决策变量称为空格决策变量，一般与线性规划某个可行解的非基变量相对应，空格决策变量的检验数简称为空格检验数。

第一列的决策变量确定了以后，这个运输问题就变成一个 3 个产地、3 个销地的运输问题，注意 A_1 的产量变为 4，这是因为已经运走了 3 个单位物质到 B_1，见表 3-9。

北京大学出版社版权所有 禁止转载

表 3-9 西北角法二

产地 \ 销地	B_2	B_3	B_4	供应量
A_1	11	3	10	4
A_2	9	2	8	4
A_3	4	10	5	9
需要量	6	5	6	17

对表 3-9 重复上面过程，找出其西北角方格 x_{12}， $x_{12}=\min\{6,4\}=4$，这时 A_1 的供应量全部调拨完毕，所以该行的决策变量调运量都已经确定，划去该行，见表 3-10，就得到一个 2 个产地、3 个销地的运输问题，注意 B_2 的需求量变为 2。

表 3-10 西北角法三

产地 \ 销地	B_2	B_3	B_4	供应量
A_2	9	2	8	4
A_3	4	10	5	9
需要量	2	5	6	13

北京大学出版社版权所有 禁止转载

一直进行下去，直至所有的物质调运完毕，就得到一个初始调运方案，分配完所有的物质，见表 3-11。上述过程旨在示范求解过程，实际求解将表 3-8 中的行列划掉即可。

表 3-11 西北角法四

产地 \ 销地	B_1	B_2	B_3	B_4	供应量
A_1	3 3	11 4	3	10	7
A_2	1	9 2	2 2	8	4
A_3	7	4	10 3	5 6	9
需要量	3	6	5	6	20

西北角法确定出的初始调运方案为

由 $A_1 \to B_1$ 运 3；$A_1 \to B_2$ 运 4；$A_2 \to B_2$ 运 2；$A_2 \to B_3$ 运 2；$A_3 \to B_3$ 运 3；$A_3 \to B_4$ 运 6。方案的运输总费用为

$$z = 3\times3+4\times11+2\times9+2\times2+3\times10+6\times5=135$$

2. 最小元素法

最小元素法是按照“最低运输成本优先集中供应”的原则，即运价最小的需求优先满足，即从单位运价中最小的运价开始确定供需关系，然后依次找单位运价的次小值，一直到给出初始基本可行解为止。最小元素法的基本思想是就近供应，每一次都要求找出单位运价表中最小的元素，在运量表内对应的方格填入允许取得的最大数，若某行(列)的供应量(需求量)已满足，则把运价表中该运价所在行(列)划去；再找出未划去的单位运价中的最小数值。一直进行下去，直至得到一个基可行解。

【例 3.3】求表 3-12 所给运输问题的初始调运方案。

表 3-12　单位运价表

销地 / 产地	B_1	B_2	B_3	B_4	供应量
A_1	12	13	10	11	7
A_2	10	12	14	10	9
A_3	14	11	15	12	7
需求量	3	5	7	8	23

解：首先在单位运价表中找出最小的运价(若几个同时为最小，则任取其中一个)，取最小单位运价 $c_{21}=10$，这表示先将 A_2 产品供应给 B_1 是最经济的，故优先满足 B_1 的需求，$x_{21}=\min\{3,9\}=3$；在表 3-13 中的 A_2B_1 方格处填上 3，这样 B_1 的需求已满足，已不需要再向它供货，故在表 3-13 中划去 B_1 所在的列；这时 A_2 的产量变为 9−3=6，接着考虑这个运输问题。一直进行下去，就得到该运输问题的初始调运方案。

表 3-13　最小元素法一

销地 / 产地	B_1	B_2	B_3	B_4	供应量
A_1	12	13	10 7	11	7
A_2	10 3	12	14	10 6	9
A_3	14	11 5	15	12 2	7
需要量	3	5	7	8	23

由表 3-13 可以看出，该初始调运方案只有 5 个决策变量的取值不为 0。根据定理 3.1，该运输问题可行基变量的个数应为 3+4−1=6，说明这里得到的是一个退化可行解，有一个基变量取值为 0。退化解会影响表上作业法——换基迭代的进行，当得到退化解时，要增加一些变量凑足初始可行基变量的个数 $m+n-1$。一般选择在某行与某列只有一个基变量的交叉处，这些变量有助于构建运输问题的闭回路。尽管这些变量的取值等于零，但仍认为该方格有数(即认为是非空格决策变量)，并将其取值记为 0，见表 3-14。在换基迭代时仍要注意这种情况的出现，如果换基同时将几个基变量的取值都变成零，一定将那些不作为

北京大学出版社版权所有 禁止转载

换基变量的决策变量仍作为新基的基变量，尽管其取值为零，但仍要标上数值 0，以保证基变量的个数不变。

表 3-14　最小元素法二

销地 产地	B_1	B_2	B_3	B_4	供应量
A_1	12 0	13	10 7	11	7
A_2	10 3	12	14	10 6	9
A_3	14	11 5	15	12 2	7
需要量	3	5	7	8	23

最小元素法确定出的初始调运方案为

由 $A_1 \to B_3$ 运 7；$A_2 \to B_1$ 运 3；$A_2 \to B_4$ 运 6；$A_3 \to B_2$ 运 5；$A_3 \to B_4$ 运 2。方案的运输总费用为

$$z = 7\times10 + 3\times10 + 6\times10 + 5\times11 + 2\times12 = 239$$

【例 3.4】用最小元素法求例 3.2 运输问题的初始调运方案。

解：采用最小元素法，确定的调运方案见表 3-15。

北京大学出版社版权所有　禁止转载

表 3-15　调运方案

销地 产地	B_1	B_2	B_3	B_4	供应量
A_1	3	11	3 4	10 3	7
A_2	1 3	9	2 1	8	4
A_3	7	4 6	10	5 3	9
需要量	3	6	5	6	20

初始调运方案为

$A_1 \to B_3$ 运 4；$A_1 \to B_4$ 运 3；$A_2 \to B_1$ 运 3；$A_2 \to B_3$ 运 1；$A_3 \to B_2$ 运 6；$A_3 \to B_4$ 运 3。其运输总费用为

$$z = 4\times3 + 3\times10 + 3\times1 + 1\times2 + 6\times4 + 3\times5 = 86$$

3. *差值法*

差值法又称为伏格尔法(Vogel 法)。最小元素法只考虑了局部的运输费用最小，有时为了节省某一处运费，可能会导致其他处运费很大，缺乏对整个供需关系的考虑。差值法考虑产地与销地最小和次小运价的差额，如果差额很大，就选最小运价先调运，不然就会增加总的费用。差值法具体包括以下几个步骤。

(1) 算出单位运价表各行各列中最小元素和次小元素的差额的绝对值。

(2) 在差额最大的行或列中的最小元素处填上尽可能大的调运数(若几个差额同为最大，则可任取其一)。

(3) 这时必有一列或一行调运完毕，在剩下的运价表中再求最大差额，进行第二次调运，直至最后调运完毕，就得到一个初始调运方案。

由此可见，差值法同最小元素法除在确定供求关系的原则上不同外，其余步骤相同。实践结果表明，由差值法给出的初始解，比用西北角法或最小元素法给出的初始解更接近最优解，参看例3.5。

【例3.5】用差值法求例3.2的初始调运方案。

解：首先求各行各列最小和次小元素的差值，见表3-16中k_i、l_j，由于$\max\limits_{i,j}\{k_i, l_j\} = 5$，选取第2列、第3行所对应最小运费所在的方格，填入最大调运数量6，划去第2列。

表3-16　调运方案

产地 \ 销地	B_1	B_2	B_3	B_4	供应量	k_i
A_1	3	11	3 / 5	10 / 2	7	0
A_2	1 / 3	9	2	8 / 1	4	1
A_3	7	4 / 6	10	5 / 3	9	1
需要量	3	6	5	6	20	
l_j	2	5	1	3		

北京大学出版社版权所有 禁止转载

重复上面计算，直至得到初始调运方案，见表3-16。得到的初始调运方案为：$A_1 \to B_3$运5；$A_1 \to B_4$运2；$A_2 \to B_1$运3；$A_2 \to B_4$运1；$A_3 \to B_2$运6；$A_3 \to B_4$运3。此时运输总费用为

$$z = 5\times3 + 2\times10 + 3\times1 + 1\times8 + 6\times4 + 3\times5 = 85$$

西北角法、最小元素法和差值法这3种确定初始调运方案的方法各有千秋，西北角法比较机械，完全按照位置来找初始可行解，对于计算机编程来说较易实现；最小元素法、差值法较为灵活，其初始可行解比较接近最优解，所以换基迭代步骤较少，便利于手工计算；但差值法的计算最为烦琐，虽说更接近最优解，但在实际问题中多用最小元素法确定初始调运方案。

3.2.2　检验数的计算

表上作业法计算检验数的方法有两种：一种是闭回路法；另一种是位势法。

1. 闭回路法

当初始调运方案确定后，就确定了运输问题的初始可行解，也就确定了基变量和非基变量。根据定理3.2，对每一个非基变量可以找到而且只能找到唯一的一个闭回路，因而在产销平衡表3-5中，从任一空格(没有调运量的格，即非基变量所在的格)出发，可找到唯一一个闭回路。

对于表 3-15 给出的初始调运方案，(A_1,B_1)为一空格。如果从(A_1,B_1)空格出发，沿(A_1,B_3)、(A_2,B_3)、(A_2,B_1) 3 个有数字的格，又回到(A_1,B_1)空格，从而形成$(A_1,B_1)\to(A_1,B_3)\to(A_2,B_3)\to(A_2,B_1)\to(A_1,B_1)$闭回路，即变量

$$x_{11}\to x_{13}\to x_{23}\to x_{21}\to x_{11}$$

组成的闭回路，以此类推，可以找出所有空格的闭回路，见表 3-17，格(A_i,B_j)简记为(i, j)。

表 3-17　闭回路

空格变量	空格位置	闭　回　路
x_{11}	(1, 1)	(1, 1)—(1, 3)—(2, 3)—(2, 1)—(1, 1)
x_{12}	(1, 2)	(1, 2)—(1, 4)—(3, 4)—(3, 2)—(1, 2)
x_{22}	(2, 2)	(2, 2)—(2, 3)—(1, 3)—(1, 4)—(3, 4)—(3, 2)—(2, 2)
x_{24}	(2, 4)	(2, 4)—(2, 3)—(1, 3)—(1, 4)—(2, 4)
x_{31}	(3, 1)	(3, 1)—(3, 4)—(1, 4)—(1, 3)—(2, 3)—(2, 1)—(3, 1)
x_{33}	(3, 3)	(3, 3)—(3, 4)—(1, 4)—(1, 3)—(3, 3)

根据单纯形法原理，要判断某个可行解是否为最优解，需要计算非基变量的检验数。用闭回路法求检验数时，对于给定的调运方案(基可行解)，从非基变量x_{ij}出发作一条闭回路，并从x_{ij}开始将该闭回路上的顶点顺序编号(顺时针或逆时针均可)，起点为 0，以此类推。称编号为奇数的点为奇点，编号为偶数的点为偶点，则x_{ij}对应的检验数σ_{ij}等于该闭回路上偶点处单位运价的总和与奇点处单位运价的总和之差，即

$$\sigma_{ij}=\text{偶点处单位运价的总和}-\text{奇点处单位运价的总和}$$

【例 3.6】求例 3.4 给出的可行解所对应的非基变量检验数。

解：在表 3-17 中，x_{12}的闭回路为$x_{12}\to x_{14}\to x_{34}\to x_{32}\to x_{12}$，以$x_{12}$为顶点编号，分别为$0\to1\to2\to3\to0$，$x_{12}$的检验数为

$$\sigma_{12}=c_{12}+c_{34}-c_{14}-c_{32}=11+5-10-4=2$$

其他非基变量检验数(也叫空格检验数)的计算结果见表 3-18 中的“(　)”里边的数值。

表 3-18　空格检验数

销　地 产　地	B_1	B_2	B_3	B_4	供　应　量
A_1	3 (1)	11 (2)	3 4	10 3	7
A_2	1 3	9 (1)	2 1	8 (−1)	4
A_3	7 (10)	4 6	10 (12)	5 3	9
需要量	3	6	5	6	20

得出所有非基变量的检验数后，就可以对该调运方案是否最优作出判断。如果全部检验数均为正数或零，则调运方案一定为最优方案；如果检验数中仍存在负数，则调运方案

北京大学出版社版权所有　禁止转载

不是最优方案。在表 3-18 中，因为检验数 $\sigma_{24}=-1$，所以该调运方案不是最优方案，需要进一步调整。

从检验数的计算公式可知，检验数是闭回路上单位运价的代数和，而单位运价是调运一个单位物质的运费。前面曾提到过闭回路是一个局部的调运方案，例如表 3-18 中 x_{12} 的闭回路，就是一个局部调运方案，见表 3-19。

表 3-19　局部调运方案

产地 \ 销地	B_2	B_4	供应量
A_1	11	10　3	3
A_3	4　6	5　3	9
需要量	6	6	

沿着 x_{12} 闭回路调整一个单位的运量，即偶点增加一个单位的运量，奇点减少一个单位的运量，得到表 3-20 的调运方案。

表 3-20　调运方案

产地 \ 销地	B_2	B_4	供应量
A_1	11　1	10　2	3
A_3	4　5	5　4	9
需要量	6	6	

北京大学出版社版权所有　禁止转载

比较表 3-19 和表 3-20 这两个调运方案，可知增加的运费恰好为

$$\sigma_{12}=c_{12}+c_{34}-c_{14}-c_{32}=11+5-10-4=2$$

这说明表 3-20 给出的调运方案不如表 3-19 给出的调运方案好，因为沿着闭回路调整一个单位的运量会增加运费 2 个单位，所以表 3-19 给出的调运方案是最优的，不需调整表 3-19 给出的调运方案，从而在这个局部说明表 3-18 给出的调运方案是最好的。

在表 3-18 中由于 $\sigma_{24}=-1$，说明沿着 x_{24} 的闭回路调整一个单位的运量，会使该调运方案的总运费减少一个单位，从而从这个局部说明表 3-18 给出的调运方案不是最好的，因而需要调整这个局部调运方案。

2. 位势法

根据对偶理论，若设 $u_1,u_2,\cdots,u_m$ 分别表示前 m 个约束等式相应的对偶变量；$v_1,v_2,\cdots,v_n$ 分别表示后 n 个等式约束相应的对偶变量，则产销平衡运输问题(3-1)的对偶问题为

$$\max w=\sum_{i=1}^{m}a_iu_i+\sum_{j=1}^{n}b_jv_j$$

$$\begin{cases} u_i + v_j \leqslant c_{ij} & i=1,2,\cdots,m;\ j=1,2,\cdots,n \\ u_i, v_j \text{为任意实数} \end{cases} \tag{3-5}$$

显然该对偶问题有可行解。

根据互为对偶问题解的关系，运输问题式(3-1)的检验数

$$\sigma_{ij} = c_{ij} - \boldsymbol{C}_B \boldsymbol{B}^{-1} \boldsymbol{P}_{ij}$$

其中 $\boldsymbol{C}_B \boldsymbol{B}^{-1} = (u_1, u_2, \cdots, u_m, v_1, v_2, \cdots, v_n)$，而每个决策变量 x_{ij} 的系数向量 $\boldsymbol{P}_{ij} = e_i + e_{m+j}$，这样 $\boldsymbol{C}_B \boldsymbol{B}^{-1} \boldsymbol{P}_{ij} = u_i + v_j$，$\sigma_{ij} = c_{ij} - \boldsymbol{C}_B \boldsymbol{B}^{-1} \boldsymbol{P}_{ij} = c_{ij} - (u_i + v_j)$。

从单纯形理论可知所有基变量的检验数等于 0。若 x_{ij} 为基变量，则

$$\sigma_{ij} = c_{ij} - (u_i + v_j) = 0 \tag{3-6}$$

从而可以构建一组方程组解出 u_i、v_j，代回到检验数 $\sigma_{ij} = c_{ij} - (u_i + v_j)$ 公式中，就可求出空格(非基变量)的检验数。

因为运输问题的可行基变量的个数为 $m+n-1$，所以可确定出检验数的 $m+n-1$ 个方程。为求出 $m+n$ 个变量对偶问题的解，需事先规定一个自由变量，一般选择 u_1 或 v_n 即可，然后令 $u_1 = 0$ 或 $v_n = 0$，就得对偶问题的一个解。这样求得的对偶问题的解 u_i^* 与 v_j^* 分别称为对应调运方案的第 i 行的“行位势”、第 j 列的“列位势”，$u_i^* + v_j^*$ 为变量 x_{ij} 的位势，可以借助位势求出非基变量的检验数。如果某个对偶问题的解使得所有 $\sigma_{ij} = c_{ij} - (u_i + v_j) \geqslant 0$，那么该解就是对偶问题的可行解，根据弱对偶定理 2.9，这时就得到原运输问题的最优解。

位势法可以在表上作业实现，其包括以下几个步骤。

(1) 在产销平衡表中，即调运方案中增加 u_i 行和 v_j 列，但在相应的基变量格(即数字格中)不是填写调运量，而是填写相应的运价，写在格的左上角。

(2) 令 $u_1 = 0$ 或 $v_n = 0$，根据式(3-6)依一定次序计算 u_i 和 v_j，将结果填写在表中。

(3) 将非基变量的运价填入相应格的左上角，根据式(3-6)计算相应的检验数 σ_{ij}，将结果填入相应格 (i, j)，写在该格的右下角。

【例 3.7】应用位势法求例 3.4 给出的初始可行解所对应的非基变量的检验数。

解： (1) 在表 3-18 上增加一行一列，在列中填入 $u_i (i=1,2,3)$，在行中填入 $v_j (j=1,2,3,4)$，得表 3-21。令 $u_1 = 0$，然后按照 $c_{ij} = u_i + v_j$，相继地确定出 u_i、v_j，填入表 3-21 中。如 $c_{13} = u_1 + v_3$，可得 $v_3 = 3$，以此类推，确定所有的 u_i、v_j。

表 3-21　空格检验数

产地＼销地	B_1	B_2	B_3	B_4	u_i
A_1	3	11	3　　4	10　　3	0
A_2	1　　3	9	2　　1	8	−1
A_3	7	4　　6	10	5　　3	−5
v_j	2	9	3	10	

(2) 利用公式 $\sigma_{ij} = c_{ij} - u_i - v_j$，计算所有的空格检验数，如

$$\sigma_{11} = c_{11} - u_1 - v_1 = 3 - 0 - 2 = 1$$

$$\sigma_{12} = c_{12} - u_1 - v_2 = 11 - 0 - 9 = 2$$

将计算结果填入表 3-22，由于 x_{24} 的检验数为负，说明该解不是最优解，还需继续改进。

表 3-22　空格检验数

销地 产地	B_1	B_2	B_3	B_4	u_i
A_1	3 (1)	11 (2)	3 4	10 3	0
A_2	1 3	9 (1)	2 1	8 (−1)	−1
A_3	7 (10)	4 6	10 (12)	5 3	−5
v_j	2	9	3	10	

闭回路法的主要缺点是当变量个数较多时，寻找闭回路以及计算两方面都会产生困难。当运输问题的产地与销地很多时，空格的数目很多，用闭回路法计算检验数，要找很多的闭回路，计算量很大，而用位势法就简便得多。

3.2.3　闭回路的调整

当初始基本可行解非基变量的检验数出现负数时，便需换基迭代，具体包括以下几个步骤。

(1) 若有两个和两个以上的负检验数时，一般选择其中最小的负检验数，以它对应的空格为调入格，以此格为出发点，作一闭回路，并从该空格出发，沿闭回路，将各顶点依次编号，空格编号为 0。

表 3-22 中只有一个负检验数 $\sigma_{24} = -1$，以空格 (A_2, B_4) 为出发点，考虑闭回路 $(A_2, B_4) \rightarrow (A_2, B_3) \rightarrow (A_1, B_3) \rightarrow (A_1, B_4) \rightarrow (A_2, B_4)$，见表 3-23。

表 3-23　闭回路

销地 产地	B_1	B_2	B_3	B_4
A_1			3 4	10 3
A_2			2 1	8 (−1)
A_3				

北京大学出版社版权所有
禁止转载

(2) 取奇点所在格中最小的运量，即令 $\theta = \min(3,1) = 1$，然后在闭回路中偶点增加 θ、奇点减少 θ，得出新的调整方案，见表 3-24。

表 3-24　新调运方案

产地 \ 销地	B_1	B_2	B_3	B_4
A_1	3	11	3 / 5	10 / 2
A_2	1 / 3	9	2	8 / 1
A_3	7	4 / 6	10	5 / 3

重新计算空格检验数，见表 3-25。如果所有的检验数都为正数或零，那么求出的就是最优解，不然，重复上述过程。

表 3-25　空格检验数

产地 \ 销地	B_1	B_2	B_3	B_4	供应量
A_1	3 / (0)	11 / (2)	3 / 5	10 / 2	7
A_2	1 / 3	9 / (2)	2 / (1)	8 / 1	4
A_3	7 / (9)	4 / 6	10 / (12)	5 / 3	9
需要量	3	6	5	6	20

北京大学出版社版权所有　禁止转载

表 3-25 中的所有检验数都非负，故表 3-25 中的解为最优解，最小总运费为

$$(5\times3+2\times10+3\times1+1\times8+6\times4+3\times50)\text{元}=85\text{元}$$

3.3　其他形式的运输问题

产销平衡运输问题相当于线性规划的标准型，实际问题中经常还会遇到一些其他的运输问题，解决的主要方法是将这些问题都转化为产销平衡的运输问题。

3.3.1　产销不平衡的运输问题

产销不平衡的运输问题分两类：一是供大于求的运输问题；二是供不应求的运输问题。

1. 供大于求的运输问题

供大于求的运输问题，即在 $\sum_{i=1}^{m}a_i > \sum_{j=1}^{n}b_j$ 的情况下，求 $\min z=\sum_{j=1}^{n}\sum_{i=1}^{m}c_{ij}x_{ij}$ (总费用最少)，得供大于求的运输问题模型为

$$\min z=\sum_{j=1}^{n}\sum_{i=1}^{m}c_{ij}x_{ij}$$

$$\begin{cases}\sum_{j=1}^{n}x_{ij}\leqslant a_i & i=1,2,\cdots,m\\ \sum_{i=1}^{m}x_{ij}=b_j & j=1,2,\cdots,n\\ x_{ij}\geqslant 0 & i=1,2,\cdots,m;\ j=1,2,\cdots,n\end{cases}\tag{3-7}$$

式(3-7)仍然是一个线性规划问题，求解时要变为线性规划的标准型，为此，将约束条件的前 m 个不等式添加松弛变量 $x_{i,n+1}\geqslant 0$，这样不等式约束就变成 $\sum_{j=1}^{n}x_{ij}+x_{i,n+1}=a_i$ 等式约束，模型就变为

$$\min z=\sum_{j=1}^{n}\sum_{i=1}^{m}c_{ij}x_{ij}$$

$$\begin{cases}\sum_{j=1}^{n}x_{ij}+x_{i,n+1}=a_i & i=1,2,\cdots,m\\ \sum_{i=1}^{m}x_{ij}=b_j & j=1,2,\cdots,n\\ x_{ij}\geqslant 0 & i=1,2,\cdots,m;\ j=1,2,\cdots,n,n+1\end{cases}$$

北京大学出版社版权所有 禁止转载

对照模型(3-1)，将前 m 行相加减去后 n 就得到 $\sum_{i=1}^{m}x_{i,n+1}=\sum_{i=1}^{m}a_i-\sum_{j=1}^{n}b_j=b_{n+1}$，因而可以添加一个冗余约束 $\sum_{i=1}^{m}x_{i,n+1}=b_{n+1}$ 就将式(3-7)变为产销平衡的运输问题

$$\min z=\sum_{j=1}^{n}\sum_{i=1}^{m}c_{ij}x_{ij}$$

$$\begin{cases}\sum_{j=1}^{n}x_{ij}+x_{i,n+1}=a_i & i=1,2,\cdots,m\\ \sum_{i=1}^{m}x_{ij}=b_j & j=1,2,\cdots,n\\ \sum_{i=1}^{m}x_{i,n+1}=b_{n+1} & \\ x_{ij}\geqslant 0 & i=1,2,\cdots,m\quad j=1,2,\cdots,n,n+1\end{cases}$$

事实上，上面模型就是 m 个产地和 $n+1$ 个销地的产销平衡运输问题，相当于增加了一个“虚拟”销地，由于该销地并不存在，每个产地的剩余物资只能留在原产地，因此产地 A_i

运到 B_{n+1} 的单位运价为 $c_{i,n+1}=0$，而该销地的销量是 b_{n+1}。因而供大于求运输问题的求解思路是添加一个虚拟销地，转化为平衡运输问题来处理。

2. 供不应求的运输问题

供不应求的运输问题，即 $\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j$ 的情况下，求 $\min z=\sum_{j=1}^{n}\sum_{i=1}^{m} c_{ij}x_{ij}$，就得线性规划模型为

$$\min z=\sum_{j=1}^{n}\sum_{i=1}^{m} c_{ij}x_{ij}$$
$$\begin{cases}\sum_{j=1}^{n} x_{ij}=a_i & i=1,2,\cdots,m \\ \sum_{i=1}^{m} x_{ij}\leqslant b_j & j=1,2,\cdots,n \\ x_{ij}\geqslant 0 & i=1,2,\cdots,m;\ j=1,2,\cdots,n\end{cases} \tag{3-8}$$

这时可加入一个虚设的产地 A_{m+1} 去生产不足的物资，相当于后面 n 个约束中都添加一个松弛变量。假设它的供给量为 $\sum_{j=1}^{n} b_j-\sum_{i=1}^{m} a_i=a_{m+1}$，以其生产的物资去弥补各销地需求量的不足；实际上各销地得不到该产地提供的物资，因而 A_{m+1} 到 B_j 的运费应为 $c_{m+1,j}=0$，从而将供不应求的运输问题转化为产销平衡的运输问题。

【例 3.8】某公司在不同地区有 A_1、A_2 和 A_3 三个工厂，产品将运往 B_1、B_2、B_3 和 B_4 四个地区销售，其单位运价表见表 3-26，试建立产销平衡的运输问题的数学模型。

表 3-26 单位运价表

销地 产地	B_1	B_2	B_3	B_4	供应量 a_i
A_1	6	10	16	8	60
A_2	14	8	16	12	100
A_3	20	6	10	4	120
需求量 b_j	30	20	80	90	220/280

北京大学出版社版权所有 禁止转载

解：这是一个供大于求的物资调运问题，故增加一个虚拟的销地 B_5，B_5 需求量为 $b_5=280-220=60$，令 $c_{i5}=0, i=1,2,3$，该供需平衡运输问题的数学模型为

$$\min z=\sum_{j=1}^{5}\sum_{i=1}^{3} c_{ij}x_{ij}$$
$$\begin{cases}\sum_{j=1}^{5} x_{ij}=a_i & i=1,2,3 \\ \sum_{i=1}^{3} x_{ij}=b_j & j=1,2,3,4,5 \\ x_{ij}\geqslant 0 & i=1,2,3;\ j=1,2,3,4,5\end{cases}$$

实际上，上述模型可以从单位运价表 3-27(a)简单明了地得出。

表 3-27(a) 单位运价表

产地 \ 销地	B_1	B_2	B_3	B_4	B_5	供应量
A_1	6	10	16	8	0	60
A_2	14	8	16	12	0	100
A_3	20	6	10	4	0	120
需求量	30	20	80	90	60	280

该问题的最优调运方案见表 3-27(b)。

表 3-27(b) 产销平衡表

产地 \ 销地	B_1	B_2	B_3	B_4	B_5	供应量
A_1	30			30		60
A_2		20	20		60	100
A_3			60	60		120
需求量	30	20	80	90	60	280

北京大学出版社版权所有 禁止转载

3.3.2 禁运与封锁的运输问题

在实际的物资运输管理中常遇到以下情况，某种物资不能从产地 A_i 运往销地 B_j，或者销地 B_j 不接收从产地 A_i 调入的物资，称前者为 A_i 对 B_j 的禁运，后者为 B_j 对 A_i 的封锁。造成禁运或封锁的因素很多，例如 A_i 与 B_j 之间没有运输线，或者由于自然灾害造成了原有交通运输线的中断，这样就形成了 A_i 对 B_j 的禁运；如果物资需通过铁路、航运运输，由于运输能力有限，有关部门暂时禁止这批物资通过他们所管辖的路段，也人为地造成了 A_i 对 B_j 的禁运；由于某种经济原因，如质量问题或合同约束，销地 B_j 拒绝接收产地 A_i 的物资，从而形成 B_j 对 A_i 的封锁。

禁运和封锁给物资运输管理工作带来的后果是在制定物资调运方案时，必须使物资从 A_i 到 B_j 的调运量为零。也就是说，在数学模型中要增加约束条件 $x_{ij}=0$，去掉这个约束条件使模型转化为运输问题的方法是：将 A_i 到 B_j 的运价 c_{ij} 修改为一个充分大的正数 M，从而使得任意一个含有 $x_{ij}\neq 0$ 的调运方案均不可能成为最优方案，这样在得到了相应的运输问题的最优调运方案时，约束条件 $x_{ij}=0$ 自动地得到了满足，与线性规划的大 M 法相对应。

【例 3.9】供需双方在协商后签订了一个供货合同，合同规定供给 6 个地区(记为 $B_1, B_2, \cdots, B_6$)提供某种物资并负责物资的运输，同时规定 B_2 和 B_4 的物资只能由产地 A_1 或 A_2 调入，各地的供给量、需求量和单位运价由表 3-28 给出，试求满足合同要求的最优调运方案。

表 3-28 单位运价表

产地 \ 销地	B_1	B_2	B_3	B_4	B_5	B_6	供应量
A_1	3	5	7	9	2	8	150
A_2	1	3	2	6	4	5	180

续表

产地＼销地	B_1	B_2	B_3	B_4	B_5	B_6	供应量
A_3	2	2	6	3	4	6	120
需求量	70	80	40	140	60	60	450

解：合同要求 B_2 和 B_4 的物资只能由 A_1 或 A_2 供给，形成了 B_2 和 B_4 对 A_3 的封锁。将 c_{32} 和 c_{34} 改为 M，见表 3-29。

表 3-29　单位运价表

产地＼销地	B_1	B_2	B_3	B_4	B_5	B_6	供应量
A_1	3	5	7	9	2	8	150
A_2	1	3	2	6	4	5	180
A_3	2	M	6	M	4	6	120
需求量	70	80	40	140	60	60	450

应用表上作业法求解该运输问题，得到最优调运方案，由表 3-30 给出。

表 3-30　最优调运方案表

产地＼销地	B_1	B_2	B_3	B_4	B_5	B_6	供应量
A_1	10	80			60		150
A_2			40	140			180
A_3	60					60	120
需求量	70	80	40	140	60	60	450

北京大学出版社版权所有 禁止转载

3.3.3　运力限制的运输问题

在制订物资调运方案时，管理人员应该考虑物资所经路段的运输能力。设 A_i 到 B_j 路段的运输能力为 d_{ij}，如果 A_i 的供应量和 B_j 的需求量都大于 d_{ij}，则从 A_i 到 B_j 的物资调运量至多为 d_{ij}。也就是说，在物资调运时 A_i 到 B_j 路段存在运输能力的限制，此时相应的数学模型中应增加运输能力约束，即有 $x_{ij} \leqslant d_{ij}$。为将这种类型的问题转化为运输问题模型，可将 B_j 想象为两个销地 B_j' 和 B_j''，规定 B_j' 的需求量为 d_{ij}，从而使得 A_i 到 B_j' (实际上为 A_i 到 B_j) 路段不再有运输能力的限制，同时规定 B_j'' 的需求量为 $b_j - d_{ij}$，且 B_j'' 对 A_i 封锁，这样就不会有多于 d_{ij} 的物资经过 A_i 到 B_j 的路段。

【例 3.10】某运输公司可承担某种物资的运输任务，有关数据由表 3-31 给出，其中 c_{ij} 表示单位物资从 A_i 到 B_j 的运价。有关部门在 A_1 到 B_3、A_2 到 B_1、A_2 到 B_4 3 个路段给出该公司的物资通过限量分别为 15，15 和 10。应如何指定物资调运方案，才能使运输的总成本最小。

表 3-31　单位运价表

产地 \ 销地	B_1	B_2	B_3	B_4	供应量
A_1	9	5	3	10	25
A_2	6	3	7	2	55
A_3	3	8	4	2	20
需求量	45	15	20	20	100

解：由于 A_2 的供应量和 B_1 的需求量都大于该路段的限制量，上述问题在 A_2 到 B_1 路段具有运输能力限制。为建立该问题的运输问题模型，将 B_1 视为两个销地 B_1' 和 B_1''，需求量分别为 15 和 30，且 B_1'' 对 A_2 封锁。同样处理另外两个有运输能力限制的路段，A_1 到 B_3 和 A_2 到 B_4，具体的处理结果见表 3-32。

表 3-32　单位运价表

产地 \ 销地	B_1'	B_1''	B_2	B_3'	B_3''	B_4	B_4''	供应量
A_1	9	9	5	3	M	10	10	25
A_2	6	M	3	7	7	2	M	55
A_3	3	3	8	4	4	2	2	20
需求量	15	30	15	15	5	10	10	100

应用表上作业法求解，得到调运方案，见表 3-33。

表 3-33　最优调运方案

产地 \ 销地	B_1'	B_1''	B_2	B_3'	B_3''	B_4	B_4''	供应量
A_1		20		5				25
A_2	15		15	10	5	10		55
A_3		10					10	20
需求量	15	30	15	15	5	10	10	100

将 B_j' 和 B_j'' 合并视为 B_j $(j=1,3,4)$，就得到可操作的调运方案，见表 3-34。

表 3-34　可行调运方案

产地 \ 销地	B_1	B_2	B_3	B_4	供应量
A_1	20		5		25
A_2	15	15	15	10	55
A_3	10			10	20
需求量	45	15	20	20	100

北京大学出版社版权所有
禁止转载

3.3.4　转运运输问题

在运输管理中，经常要处理物资中转的运输问题，例如，物资从产地运到销地必须使用不同的运输工具，这样就需要首先将物资从产地运到某地(称为中转站)，更换运输工具后再运往销地。又如，由于运输能力的限制或价格因素(转运运价小于直接运价)，需要将不同产地的物资首先集中到某个中转站，再由中转站发往销地。

需要中转站的运输称为转运运输。这里讨论一次转运问题，一般提法是：设有 r 个中转站 $T_1, T_2, \cdots, T_r$，物资的运输过程是先从产地 A_i 运到某个中转站 T_k，再运往销地 B_j。已知 A_i 到 T_k 的运价为 c_{ik}，T_k 到 B_j 的运价为 c_{kj}，A_i 的供给量为 a_i，通过 T_k 的最大运输能力为 d_k，B_j 的需求量为 b_j。不妨设 $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j, \sum_{i=1}^{m} a_i \leqslant \sum_{k=1}^{r} d_k$，也就是说，供需是平衡的且所有的物资经转运后都能送达销地。现在要求一转运方案，使得运输的总费用最小。

为建立转运问题的模型，设决策变量如下。

x_{ik}：从 A_i 到 T_k 的调运量，$i = 1, 2, \cdots, m; k = 1, 2, \cdots, r$

x_{kj}：从 T_k 到 B_j 的调运量，$k = 1, 2, \cdots, r; j = 1, 2, \cdots, n$

每个变量均非负。目标函数为两个阶段费用之和达到最小，即

$$\min z = \sum_{i=1}^{m} \sum_{k=1}^{r} c_{ik} x_{ik} + \sum_{j=1}^{n} \sum_{k=1}^{r} c_{kj} x_{kj}$$

约束条件分为以下几组。

供给约束：

$$\sum_{k=1}^{r} x_{ik} = a_i \quad i = 1, 2, \cdots, m$$

运输能力约束：

$$\sum_{i=1}^{m} x_{ik} = d_k \quad k = 1, 2, \cdots, r$$

中转站平衡：

$$\sum_{i=1}^{m} x_{ik} = \sum_{j=1}^{n} x_{kj} \quad k = 1, 2, \cdots, r$$

需求约束：

$$\sum_{k=1}^{r} x_{kj} = b_j \quad j = 1, 2, \cdots, n$$

根据转运问题模型的结构可以看出，转运问题也可转化为运输问题模型求解，其方法是将每个中转站 T_k 既看成产地也看成销地，从而形成一个有 $m + r$ 个产地，有 $r + n$ 个销地的运输问题。由于物资不能由 A_i 直接到达 B_j，故 B_j 对 A_i 封锁。同样，不同的中转站之间也互相封锁。T_k 的供给量和需求量均为 d_k，从而总供给量为 $\sum_{i=1}^{m} a_i + \sum_{k=1}^{r} d_k$，总需求量为 $\sum_{j=1}^{n} b_j + \sum_{k=1}^{r} d_k$，以实现供需平衡。

北京大学出版社版权所有
禁止转载

【例 3.11】将某种物资从 A_1、A_2 和 A_3 运往 B_1、B_2、B_3，B_4 和 B_5 5 个销地，物资必须经过 T_1、T_2、T_3 和 T_4 中的任意一个中转站转运，有关数据见表 3-35 和表 3-36。试求解该转运问题。

表 3-35　产地到中转站单位运价表

中转站 / 产地	T_1	T_2	T_3	T_4	供 应 量
A_1	4	5	7	6	70
A_2	7	12	10	11	80
A_3	6	11	8	9	90
中转站的运输能力	60	90	120	70	340/240

表 3-36　中转站到销地单位运价表

中转站 / 销地	B_1	B_2	B_3	B_4	B_5
T_1	4	4	3	7	3
T_2	8	3	4	9	11
T_3	4	3	4	7	7
T_4	6	2	2	8	8
销地需求	30	20	80	50	60

该转运问题可转化为具体 7 个产地和 9 个销地的运输问题，有关数据见表 3-37，应用表上作业法可得最优调运方案，见表 3-38。

表 3-37　转运问题的单位运价表

销 地 / 产 地	T_1	T_2	T_3	T_4	B_1	B_2	B_3	B_4	B_5	供 应 量
A_1	4	5	7	6	M	M	M	M	M	70
A_2	7	12	10	11	M	M	M	M	M	80
A_3	6	11	8	9	M	M	M	M	M	90
T_1	0	M	M	M	4	4	3	7	3	60
T_2	M	0	M	M	8	3	4	9	11	90
T_3	M	M	0	M	4	3	4	7	7	120
T_4	M	M	M	0	6	2	2	8	8	70
需求量	60	90	120	70	30	20	80	50	60	580

表 3-38　最优调运方案

中转站 / 产 地	T_1	T_2	T_3	T_4	B_1	B_2	B_3	B_4	B_5	供 应 量
A_1		10		60						70
A_2	60		20							80

北京大学出版社版权所有 禁止转载

续表

产地 \ 中转站	T_1	T_2	T_3	T_4	B_1	B_2	B_3	B_4	B_5	供应量
A_3			80	10						90
T_1									60	60
T_2		80					10			90
T_3			20		30	20		50		120
T_4							70			70
需求量	60	90	120	70	30	20	80	50	60	

实际运输过程中物资可能需要多次中转，这种类型的转运问题比较复杂，但将它转化为运输问题模型时，与一次转运问题的处理思路一样，依照禁运与封锁原则，即当物资不能从一个产地或中转站直接到达另一个中转站或销地时，就应当对这两地实行禁运与封锁。

3.4　指派问题

指派问题也称分配或配置问题，是资源合理配置或最优匹配问题。指派问题通常划分为标准和非标准的指派问题。

3.4.1　指派问题的引入

在现实生活中，有各种性质的指派问题。例如，有若干项工作需要分配给若干人(或部门)来完成；有若干项合同需要选择若干个投标者来承包；有若干班级需要安排在若干教室上课等。诸如此类的问题，它们的基本要求是在满足特定的指派要求条件下，使指派方案的总体效果最佳。

【例 3.12】某厂拟派 4 个维修小组去维修 4 台机车，他们相应地完成工作所需时间 $c_{ij}\ (i, j = 1, 2, 3, 4)$ 由表 3-39 给出。如何给每个小组安排工作才能使完成任务的总时间最少？

表 3-39　维修工作时间表　　单位：天

小组 \ 机车	1	2	3	4
1	2	15	13	4
2	10	4	14	15
3	9	14	16	13
4	7	8	11	9

解：该指派问题是安排维修小组去维修机车，其决策变量为

$$x_{ij} = \begin{cases} 1, & \text{指派第 } i \text{ 个小组维修第 } j \text{ 台机床} \\ 0, & \text{不指派第 } i \text{ 个小组维修第 } j \text{ 台机床} \end{cases} \quad i, j = 1, 2, 3, 4$$

由于第 j 台机车必须且只需有一个小组去维修，得任务约束为

$$\sum_{i=1}^{4} x_{ij} = 1, \quad j = 1, 2, 3, 4$$

北京大学出版社版权所有　禁止转载

同样，第 i 个小组能而且只能维修一台机车，得人员约束为

$$\sum_{j=1}^{4} x_{ij} = 1, \quad i = 1, 2, 3, 4$$

目标函数为完成任务总时间最小，得该问题的数学模型为

$$\min z = \sum_{j=1}^{4}\sum_{i=1}^{4} c_{ij}x_{ij} = 2x_{11} + 15x_{12} + \cdots$$

$$\begin{cases} \sum_{j=1}^{4} x_{ij} = 1 & i = 1, 2, 3, 4 \\ \sum_{i=1}^{4} x_{ij} = 1 & j = 1, 2, 3, 4 \\ x_{ij} = 0, 1 & i, j = 1, 2, 3, 4 \end{cases}$$

【例 3.13】某商业公司计划开办 5 家新商店。为了尽早建成营业，商业公司决定由 5 家建筑公司分别承建。已知建筑公司 A_i $(i = 1, 2, 3, 4, 5)$ 对新商店 B_j $(j = 1, 2, 3, 4, 5)$ 的建造费用的报价(万元)为 c_{ij} $(i, j = 1, 2, 3, 4, 5)$ ，见表 3-40。商业公司应当对 5 家建筑公司怎样分派建筑任务，才能使总的建筑费用最少？

表 3-40　建造费用报价　　单位：万元

建筑公司 \ 商店	B_1	B_2	B_3	B_4	B_5
A_1	4	8	7	15	12
A_2	7	9	17	14	10
A_3	6	9	12	8	7
A_4	6	7	14	6	10
A_5	6	9	12	10	6

北京大学出版社版权所有 禁止转载

解：该指派问题是安排建筑公司去承建商店，其决策变量为

$$x_{ij} = \begin{cases} 1, & \text{指派 } A_i \text{ 承建商店 } B_j \\ 0, & \text{不指派 } A_i \text{ 承建商店 } B_j \end{cases} \quad i, j = 1, 2, 3, 4, 5$$

则该问题的数学模型为

$$\min z = \sum_{j=1}^{4}\sum_{i=1}^{4} c_{ij}x_{ij} = 4x_{11} + 8x_{12} + \cdots$$

$$\begin{cases} \sum_{j=1}^{5} x_{ij} = 1 & i = 1, 2, 3, 4, 5 \\ \sum_{i=1}^{5} x_{ij} = 1 & j = 1, 2, 3, 4, 5 \\ x_{ij} = 0, 1 & i, j = 1, 2, 3, 4, 5 \end{cases}$$

显然指派问题与运输问题相类似，该问题的指派平衡表见表 3-41。

表 3-41　指派平衡表

建筑公司＼商店	B_1	B_2	B_3	B_4	B_5	任务
A_1	(4) x_{11}	(8) x_{12}	(7) x_{13}	(15) x_{14}	(12) x_{15}	1
A_2	(7) x_{21}	(9) x_{22}	(17) x_{23}	(14) x_{24}	(10) x_{25}	1
A_3	(6) x_{31}	(9) x_{32}	(12) x_{33}	(8) x_{34}	(7) x_{35}	1
A_4	(6) x_{41}	(7) x_{42}	(14) x_{43}	(6) x_{44}	(10) x_{45}	1
A_5	(6) x_{51}	(9) x_{52}	(12) x_{53}	(10) x_{54}	(6) x_{55}	1
公司数	1	1	1	1	1	5

3.4.2　标准指派问题的数学模型

指派问题的一般提法是：有 n 项任务，需分配给 n 个人员(或设备)去完成，已知每个人员完成某项工作的效率(或成本等)为 c_{ij}，如何给每个人员指派一项工作，使得完成任务的总效率最高(或总成本最少)，见表 3-42。

表 3-42　指派问题

人员＼任务	1	2	…	n
1	c_{11}	c_{12}	…	c_{1n}
2	c_{21}	c_{22}	…	c_{2n}
⋮	⋮	⋮	⋮	⋮
n	c_{n1}	c_{n2}	…	c_{nn}

北京大学出版社版权所有 禁止转载

为了建立标准指派问题的数学模型，引入 n^2 个 0−1 变量：

$$x_{ij}=\begin{cases}1, & \text{指派第 } i \text{ 个人做第 } j \text{ 项工作}\\ 0, & \text{不指派第 } i \text{ 个人做第 } j \text{ 项工作}\end{cases}\quad i,j=1,2,\cdots,n$$

这样，该指派问题的数学模型可写成

$$\min z=\sum_{j=1}^{n}\sum_{i=1}^{n}c_{ij}x_{ij}$$

$$\begin{cases}\sum_{j=1}^{n}x_{ij}=1 & i=1,2,\cdots,n \quad (3-9)\\ \sum_{i=1}^{n}x_{ij}=1 & j=1,2,\cdots,n \quad (3-10)\\ x_{ij}=0,1 & i,j=1,2,\cdots,n\end{cases}$$

其中：式(3-9)表示每件事必有且只有一个人去做，式(3-10)表示每个人必做且只做一件事。指派问题是产量(a_i)、销量(b_j)相等，且 $a_i=b_j=1$，$i,j=1,2,\cdots,n$ 的运输问题。

表 3-42 中的数据可用一个矩阵 $\boldsymbol{C}$ 表示，称为效率矩阵(在其他问题中，可根据实际意义称为费用矩阵等)，其元素 c_{ij} 体现了第 i 个人完成第 j 项工作时的效率，即

$$C=[c_{ij}]_{n\times n}=\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

决策变量 x_{ij} 排成的 $n\times n$ 矩阵为

$$X=[x_{ij}]_{n\times n}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix}$$

称为决策变量矩阵。

指派问题解的特征是它有 n 个 1，其他都是 0，即这 n 个 1 位于决策变量矩阵的不同行、不同列，每一种情况为指派问题的一个可行解，共有 $n!$ 个解。指派问题是：把这 n 个 1 放到 X 的 n^2 个位置的什么地方可使耗费的总资源最少(解最优)?

【例 3.14】对于效率矩阵 $C=\begin{bmatrix} 5 & 0 & 2 & 0 \\ 2 & 3 & 0 & 0 \\ 0 & 5 & 6 & 7 \\ 4 & 8 & 0 & 0 \end{bmatrix}$，决策变量矩阵：

$$X^{(1)}=\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \qquad X^{(2)}=\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

都是指派问题的最优解。

北京大学出版社版权所有 禁止转载

3.4.3 指派问题的求解

指派问题既是一类特殊的整数规划问题，又是特殊的运输问题，因此可以用多种相应的解法来求解，然而这些解法都没有充分利用指派问题的特殊性质，有效地减少计算量，直到 1955 年库恩(W. W. Kuhn)提出的匈牙利法才有效地解决了指派问题。

1. 匈牙利法的理论基础

定义 3.2 独立零元素组：在效率矩阵 C 中，有一组在不同行不同列的零元素，称为独立零元素组，其每个元素称为独立零元素。

【例 3.15】已知效率矩阵 $C=\begin{bmatrix} 5 & 0 & 2 & 0 \\ 2 & 3 & 0 & 0 \\ 0 & 5 & 6 & 7 \\ 4 & 8 & 0 & 0 \end{bmatrix}$，求其独立零元素组。

解：可行解 $\{c_{12}=0,\ c_{24}=0,\ c_{31}=0,\ c_{43}=0\}$ 是一个独立零元素组，$c_{12}=0$，$c_{24}=0$，$c_{31}=0$，$c_{43}=0$ 分别称为独立零元素；$\{c_{12}=0,\ c_{23}=0,\ c_{31}=0,\ c_{44}=0\}$ 也是一个独立零元素组；而 $\{c_{14}=0,\ c_{23}=0,\ c_{31}=0,\ c_{44}=0\}$ 就不是一个独立零元素组，因为 $c_{14}=0$ 与 $c_{44}=0$ 这两个零元素位于同一列。

根据上述对效率矩阵中零元素的分析，将效率矩阵 $\boldsymbol{C}$ 中出现的独立零元素组中零元素所处的位置，在决策变量矩阵中令相应的 $x_{ij}=1$，其余的 $x_{ij}=0$，就可找到指派问题的一个最优解。

例 3.15 的解 $\boldsymbol{X}^{(1)}=\begin{bmatrix}0&1&0&0\\0&0&0&1\\1&0&0&0\\0&0&1&0\end{bmatrix}$，$\boldsymbol{X}^{(2)}=\begin{bmatrix}0&1&0&0\\0&0&1&0\\1&0&0&0\\0&0&0&1\end{bmatrix}$都是其最优解。

但是在有的问题中效率矩阵 $\boldsymbol{C}$ 中独立零元素的个数不够 n 个，这样就无法求出最优指派方案，需作进一步的分析，首先给出下述定理。

定理 3.3 设指派问题的效率矩阵为 $\boldsymbol{C}=[c_{ij}]_{n\times n}$，若将该矩阵的某一行(或某一列)的各个元素都减去同一常数 k (k 可正可负)，得到新的效率矩阵 $\boldsymbol{C}'=[c'_{ij}]_{n\times n}$，则以 $\boldsymbol{C}'$ 为效率矩阵的新的指派问题与原指派问题的最优解相同。

推论 若将指派问题的效率矩阵每一行或每一列分别减去各行或各列的最小元素，则得到新指派问题与原指派问题有相同的最优解。

定理 3.4 效率矩阵 $\boldsymbol{C}$ 中独立零元素的最多个数等于能覆盖所有零元素的最少直线数。

2. *匈牙利法求解步骤*

基于上面的定理，匈牙利法可分为以下 4 个步骤。

1) 第 1 步：变换效率矩阵，将各行各列都减去当前各行、各列中最小元素

若第 i 行只有一个零元素 $c_{ij}=0$，表示第 i 人干第 j 件工作效率最好，因此优先指派第 i 人干第 j 件工作。

【例 3.16】变换效率矩阵 $\boldsymbol{C}=[c_{ij}]$。

$$[c_{ij}]=\begin{bmatrix}2&15&13&4\\10&4&14&15\\9&14&16&13\\7&8&11&9\end{bmatrix}\begin{matrix}2\\4\\9\\7\end{matrix}\rightarrow\begin{bmatrix}0&13&11&2\\6&0&10&11\\0&5&7&4\\0&1&4&2\end{bmatrix}\rightarrow\begin{bmatrix}0&13&7&0\\6&0&6&9\\0&5&3&2\\0&1&0&0\end{bmatrix}=[b_{ij}]$$
$$\begin{matrix}0&0&4&2\end{matrix}$$

这样得到的新矩阵 $[b_{ij}]$ 中，每行每列都必然出现零元素。

2) 第 2 步：标记新矩阵的独立零元素

(1) 进行行检验。对变换后的效率矩阵进行逐行检验，若某行只有一个未标记的零元素时，用“*”将该零元素做标记，然后将被标记的零元素所在的列的其他未标记的零元素用“×”将该零元素标记。

如 $[b_{ij}]$ 中第 3 行只有一个未标记的零元素，用“*”分别将它们做标记，然后用“×”标记第 1 列其他未被标记的零元素。重复行检验，直到每一行都没有未被标记的零元素或至少有两个未被标记的零元素为止。

(2) 进行列检验。与行检验过程类似，对进行了行检验的矩阵逐列进行检验，对每列只有一个未被标记的零元素，用“*”将该零元素做一标记，然后将该元素所在行的其他未被标记的零元素用“×”将该零元素标记。重复上述列检验，直到每一列都没有未被标记的零元素或有两个未被标记的零元素为止。

北京大学出版社版权所有 禁止转载

这时可能出现以下 3 种情况：①每一行均有标记“*”出现，“*”的个数 m 恰好等于 n；②存在未标记的零元素，但它们所在的行和列中，未标记过的零元素均至少有两个；③不存在未被标记过的零元素，“*”的个数 $m<n$。

(3) 进行试指派。若情况①出现，则可进行试指派：令“*”记号的决策变量取值为 1，其他决策变量取值均为零，得到一个最优指派方案，停止计算。

例 3.16 中的 $[b_{ij}]$ 标记后，出现了情况①，可令 $x_{14}=1$，$x_{22}=1$，$x_{31}=1$，$x_{43}=1$，其余 $x_{ij}=0$。即为最优指派。

$$[b_{ij}]=\begin{bmatrix} 0^{\times} & 13 & 7 & 0^{*} \\ 6 & 0^{*} & 6 & 9 \\ 0^{*} & 5 & 3 & 2 \\ 0^{\times} & 1 & 0^{*} & 0^{\times} \end{bmatrix} \qquad \boldsymbol{X}=\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

若情况②出现，则对每行、每列的其他未被标记的零元素任选一个，加上标记“*”，即给该零元素标记“*”，然后给同行、同列的其他未被标记的零元素加标记“×”，然后再进行行、列检验，可能出现情况①或③，出现情况①就会得到最优指派，停止计算。

若情况③出现，则要转入下一步。

3) 第 3 步：作最少直线覆盖当前所有零元素

考虑例 3.13 指派问题，其效率矩阵为 $\boldsymbol{C}$，先将各行元素分别减去本行的最小元素，然后对各列也如此，即 $\boldsymbol{C}'$ 中各行各列都已出现零元素。

$$\boldsymbol{C}=\begin{bmatrix} 4 & 8 & 7 & 15 & 12 \\ 7 & 9 & 17 & 14 & 10 \\ 6 & 9 & 12 & 8 & 7 \\ 6 & 7 & 14 & 6 & 10 \\ 6 & 9 & 12 & 10 & 6 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 4 & 3 & 11 & 8 \\ 0 & 2 & 10 & 7 & 3 \\ 0 & 3 & 6 & 2 & 1 \\ 0 & 1 & 8 & 0 & 4 \\ 0 & 3 & 6 & 4 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 3 & 0 & 11 & 8 \\ 0 & 1 & 7 & 7 & 3 \\ 0 & 2 & 3 & 2 & 1 \\ 0 & 0 & 5 & 0 & 4 \\ 0 & 2 & 3 & 4 & 0 \end{bmatrix}=\boldsymbol{C}'$$

为了确定 $\boldsymbol{C}'$ 中的独立零元素，对 $\boldsymbol{C}'$ 中的零元素进行标记，即

$$\begin{array}{c} \qquad\quad \surd \\ \boldsymbol{C}'=\begin{bmatrix} 0^{\times} & 3 & 0^{*} & 11 & 8 \\ 0^{*} & 1 & 7 & 7 & 3 \\ 0^{\times} & 2 & 3 & 2 & 1 \\ 0^{\times} & 0^{*} & 5 & 0^{\times} & 4 \\ 0^{\times} & 2 & 3 & 4 & 0^{*} \end{bmatrix} \begin{array}{c} \\ \surd \\ \surd \\ \\ \\ \end{array} \end{array} \qquad \begin{bmatrix} 0 & 3 & 0 & 11 & 8 \\ 0 & 1 & 7 & 7 & 3 \\ 0 & 2 & 3 & 2 & 1 \\ 0 & 0 & 5 & 0 & 4 \\ 0 & 2 & 3 & 4 & 0 \end{bmatrix}$$

由于只有 4 个独立零元素，少于系数矩阵阶数 5，不能进行指派。为了增加独立零元素的个数，需要对矩阵作进一步的变换，首先确定覆盖当前零元素的最小直线数。

(1) 对 $\boldsymbol{C}'$ 中所有不含“*”元素的行打 √，如第 3 行。

(2) 对打 √ 的行中，所有零元素所在的列打 √，如第 1 列。

(3) 对所有打 √ 列中标记“*”元素所在的行打 √，如第 2 行。

(4) 重复上述(2)，(3)步，直到不能进一步打 √ 为止。

(5) 对未打 √ 的每一行划一直线，如第 1、4、5 行。对已打 √ 的每一列划一纵线，如第 1 列，即得到覆盖当前 0 元素的最少直线数。

北京大学出版社版权所有 禁止转载

4) 第 4 步：对矩阵未被直线覆盖过的元素中找最小元素，将打 √ 行的各元素减去这个最小元素，将打 √ 列的各元素加上这个最小元素(以避免打 √ 行中出现负元素)，这样就增加了零元素的个数，返回第 2 步

如 $\boldsymbol{C}'$ 中未被直线覆盖过的元素中，最小元素为 $c_{22}=c_{35}=1$，对打 √ 的第 2、3 行各元素都减去 1，对打 √ 的第 1 列各元素都加上 1，得到矩阵 $\boldsymbol{C}''$，返回到第二步。

$$\boldsymbol{C}'=\begin{bmatrix} 0 & 3 & 0 & 11 & 8 \\ 0 & 1 & 7 & 7 & 3 \\ 0 & 2 & 3 & 2 & 1 \\ 0 & 0 & 5 & 0 & 4 \\ 0 & 2 & 3 & 4 & 0 \end{bmatrix} \qquad \boldsymbol{C}''=\begin{bmatrix} 1 & 3 & 0 & 11 & 8 \\ 0 & 0 & 6 & 6 & 2 \\ 0 & 1 & 2 & 1 & 0 \\ 1 & 0 & 5 & 0 & 4 \\ 1 & 2 & 3 & 4 & 0 \end{bmatrix}$$

对已增加了零元素的矩阵，再标记找出独立零元素组，一直进行下去。

$$\boldsymbol{C}''=\begin{bmatrix} 1 & 3 & 0^{*} & 11 & 8 \\ 0^{\times} & 0^{*} & 6 & 6 & 2 \\ 0^{*} & 1 & 2 & 1 & 0^{\times} \\ 1 & 0^{\times} & 5 & 0^{*} & 4 \\ 1 & 2 & 3 & 4 & 0^{*} \end{bmatrix} \qquad \boldsymbol{X}=\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$\boldsymbol{C}''$ 中已有 5 个独立零元素，故可得到指派问题的最优解 $\boldsymbol{X}$，这样安排能使总的建造费用最少，建造费用为 z =(7+9+6+6+6)万元=34 万元。

北京大学出版社版权所有 禁止转载

3.4.4　非标准指派问题

在实际应用中，常会遇到非标准形式，如求最大值，人数与工作数不相等以及不可接受的配置(某人不可完成某项任务)等特殊指派问题。解决的思路是先化成标准形式，然后再用匈牙利法求解，即对效率矩阵通过适当变换使得满足匈牙利算法的条件再求解。

1. 最大化的指派问题

最大化指派问题的一般形式为

$$\max z=\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}x_{ij}$$

$$\begin{cases} \sum\limits_{i=1}^{n}x_{ij}=1 & j=1,2,\cdots,n \\ \sum\limits_{j=1}^{n}x_{ij}=1 & i=1,2,\cdots,n \\ x_{ij}=0,1 & i,j=1,2,\cdots,n \end{cases}$$

解决办法：设最大化的指派问题的系数矩阵为 $\boldsymbol{C}=[c_{ij}]_{n\times n}$，$M=\max\{c_{11},c_{12},\cdots,c_{nn}\}$，令 $\boldsymbol{B}=[b_{ij}]_{n\times n}=[M-c_{ij}]_{n\times n}$，则以 $\boldsymbol{B}$ 为效率矩阵的最小化指派问题和以 $\boldsymbol{C}$ 为效率矩阵的原最大化指派问题有相同的最优解。

【例 3.17】某工厂有 4 名工人 A_1、A_2、A_3、A_4，分别操作 4 台车床 B_1、B_2、B_3、B_4。每小时单产量见表 3-43，求产值最大的分配方案。

表 3-43　产量表

车间 / 工人	B_1	B_2	B_3	B_4
A_1	10	9	8	7
A_2	3	4	5	6
A_3	2	1	1	2
A_4	4	3	5	6

解：令 $\boldsymbol{C}=[c_{ij}]_{n\times n}=\begin{bmatrix}10&9&8&7\\3&4&5&6\\2&1&1&2\\4&3&5&6\end{bmatrix}$，$M=\max\{10,9,\cdots,6\}=10$

$$\boldsymbol{B}=[10-c_{ij}]_{n\times n}=\begin{bmatrix}0&1&2&3\\7&6&5&4\\8&9&9&8\\6&7&5&4\end{bmatrix}\to\begin{bmatrix}0&1&2&3\\3&2&1&0\\0&1&1&0\\2&3&1&0\end{bmatrix}\to\begin{bmatrix}0&0&1&3\\3&1&0&0\\0&0&0&0\\2&2&0&0\end{bmatrix}=\boldsymbol{B}'$$

$\boldsymbol{B}'$ 中有 4 个独立零元素，所以

$$\boldsymbol{X}=\begin{bmatrix}1&0&0&0\\0&0&0&1\\0&1&0&0\\0&0&1&0\end{bmatrix}$$

为最优解，最大产值为 z =(10+6+1+5)万元=22 万元。

2. 人数和事数不等的指派问题

若人数小于事数，添一些虚拟的“人”，此时这些虚拟的“人”做各件事的费用系数取为 0，理解为这些费用实际上不会发生；若人数大于事数，添一些虚拟的“事”，此时这些虚拟的“事”被各个人做的费用系数同样也取为 0。

【例 3.18】现有 4 个人，5 件工作，每人做每件工作所耗的时间见表 3-44。

表 3-44　耗时表

工作 / 工人	B_1	B_2	B_3	B_4	B_5
A_1	10	11	4	2	8
A_2	7	11	10	14	12
A_3	5	6	9	12	14
A_4	13	15	11	10	7

问指派哪个人去完成哪项工作，可使总耗时最小？

北京大学出版社版权所有 禁止转载

解：添加虚拟人 A_5，构造耗时矩阵为

$$C=\begin{bmatrix}10&11&4&2&8\\7&11&10&14&12\\5&6&9&12&14\\13&15&11&10&7\\0&0&0&0&0\end{bmatrix}\quad X=\begin{bmatrix}0&0&0&1&0\\1&0&0&0&0\\0&1&0&0&0\\0&0&0&0&1\\0&0&1&0&0\end{bmatrix}$$

应用匈牙利法求解，得到最优解 $\boldsymbol{X}$，最少耗时为 z =2+7+6+7=22。

3. 一个人可做几件事的指派问题

若某人可做几件事，则可将该人化作相同的几个"人"来接受指派。这几个"人"做同一件事的费用系数当然一样。

【例 3.19】对例 3.13 中的指派问题，为了保证工程质量，经研究决定，舍弃建筑公司 A_4 和 A_5，让技术力量较强的建筑公司 A_1、A_2、A_3 来承建 5 家商店，其投标费用见表 3-45。根据实际情况，允许每家建筑公司承建一家或两家商店，求使总费用最少的指派方案。

表 3-45 投标费用表

商店 建筑公司	B_1	B_2	B_3	B_4	B_5
A_1	4	8	7	15	12
A_2	7	9	17	14	10
A_3	6	9	12	8	7

北京大学出版社版权所有 禁止转载

解：由于每家建筑公司最多可承建两家新商店，因此，把每家建筑公司化作相同的两家建筑公司(A_i 和 A_i'，$i=1,2,3$)，因而费用矩阵变为

$$C=\begin{bmatrix}4&8&7&15&12\\4&8&7&15&12\\7&9&17&14&10\\7&9&17&14&10\\6&9&12&12&7\\6&9&12&12&7\end{bmatrix}$$

上面的系数矩阵有 6 行 5 列，为了使"人"和"事"的数目相同，引入一件虚拟"事"，使之成为标准的指派问题，其效率矩阵为

$$C'=\begin{bmatrix}4&8&7&15&12&0\\4&8&7&15&12&0\\7&9&17&14&10&0\\7&9&17&14&10&0\\6&9&12&8&7&0\\6&9&12&8&7&0\end{bmatrix}\quad X=\begin{bmatrix}0&0&1&0&0&0\\1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&0&0&0&1\\0&0&0&0&1&0\\0&0&0&1&0&0\end{bmatrix}$$

应用匈牙利法求解，得最优解 $\boldsymbol{X}$，总费用为 z =(7+4+9+7+8)万元=35 万元。

4. 某事不能由某人去做的指派问题

某事不能由某人去做，可将此人做此事的费用取作足够大的 M 。

【例 3.20】分配甲、乙、丙、丁 4 个人去完成 A、B、C、D、E 5 项任务，每人完成各项任务的时间见表 3-46。由于任务重、人数少，考虑：任务 E 必须完成，其他 4 项任务可选 3 项完成，但甲不能做 A 项工作，试确定最优分配方案，使完成任务的总时间最少。

表 3-46 耗时表

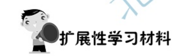
工作 / 工人	*A*	*B*	*C*	*D*	*E*
甲	25	29	31	42	37
乙	39	38	26	20	33
丙	34	27	28	40	32
丁	24	42	36	23	45

解：这是一人数与工作不等的指派问题，由于任务数大于人数，所以需要有一个虚拟的“人”，设为戊。因为甲不能做 A，所以令甲完成工作 A 的时间为 M；又因为工作 E 必须完成，故设戊完成 E 的时间为 M，即戊不能做工作 E，其余的假想时间为 0，建立的效率矩阵为 $\boldsymbol{C}$。

$$\boldsymbol{C}=\begin{bmatrix} M & 29 & 31 & 42 & 37 \\ 39 & 38 & 26 & 20 & 33 \\ 34 & 27 & 28 & 40 & 32 \\ 24 & 42 & 36 & 23 & 45 \\ 0 & 0 & 0 & 0 & M \end{bmatrix} \qquad \boldsymbol{X}=\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

北京大学出版社版权所有 禁止转载

采用匈牙利法求得最优解 $\boldsymbol{X}$，最少的耗时数 $z=(29+20+32+24)$小时=105 小时。



扩展性学习材料

协同运输管理

协同运输管理(Collaborative Transportation Management，CTM)是一种在原有“供给商-销售商”的合作关系上，扩展至“供给商—发货人—第三方物流—收货商”的战略联盟，通过信息共享和供给链协作，制订计划、猜测、运输、库存等商品服务全过程的共同决策。

协同运输是近几年发展起来的现代物流模式，它使相对独立的运输企业协同运作，以达到减少车辆空载、降低运输成本和提高企业效率等目的。在传统运输模式下，各个运输实体独立完成自身运输任务，这将不可避免地产生大量的车辆空载行驶，从而导致运输成本增加。而多个运输企业实施协同运输，不同公司之间可以互相代为完成运输任务，则可以大量减少总体的运输空载。协同运输问题在小运量、车辆可混载的情况下属于多车场装卸运输问题(Multi-depot Pickup and Delivery Problem)；在车辆限定满载情况下，为多车场弧覆盖问题(Multi-depot Arc Routing Problem)。

2000 年，全球最大零售商沃尔玛(WAL-MART)向供给商宝洁(P&G)、货运巨头亨特提出了一个新型的合作方案，要在三者间实现更透明的信息交换，共同进行决策，这就是协同运输管理的开始。他们达成

合作关系以后，沃尔玛大大减少了货物处理过程的步骤，而亨特减小了 16%的装卸货等待时间，空载率下降 3%，宝洁也实现了库存的下降。

本 章 小 结

运输和指派问题都是特殊的线性规划问题。运输问题考虑的是从出发地运送货物到目的地的调运方案的费用优化；指派问题分析的是一系列需要分配的任务与执行这些任务的被指派者的合理匹配。本章学习了运输问题的表上作业法和指派问题的匈牙利方法，并就其在物流活动中的应用做了详细的探讨。



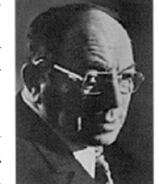
关键术语(中英文)

运输问题(Transportation Problem)　　表上作业法(Table Dispatching Method)
最小元素法(Matrix Minimum)　　闭回路法(Loop Method)
位势法(Potential Method)　　指派问题(Assignment Problem)



计划经济学家——康托洛维奇

利奥尼德·康托洛维奇(L.V.Kantorovich，1912—1986)，苏联数学家，出生于俄国圣彼得堡的一个医生家庭。1930 年毕业于列宁格勒大学，1934 年成为该校最年轻的数学教授，1935 年获该校数学博士学位，1948—1960 年任列宁格勒科学院数学所研究室主任，1958 年当选为苏联科学院通讯院士，并于 1964 年成为苏联科学院院士。1960—1971 年任苏联科学院西伯利亚分院数学所副所长，1971—1976 年任苏联国家科学技术委员会管理研究所室主任。1976 年任苏联科学院系统分析所所长。他曾于 1949 年获斯大林数学奖，1965 年获列宁经济学奖。康托洛维奇对经济学的贡献主要在于，他建立和发展了线性规划方法，并运用于经济分析，对现代经济应用数学的重要分支——线性规划方法的建立和发展作出了开创性贡献。他把资源最优利用这一传统的经济学问题，由定性研究和一般的定量分析推进到现实计量阶段，对于在企业范围内如何科学地组织生产和在国民经济范围内怎样最优地利用资源等问题做出了独创性的研究。康托洛维奇的主要著作包括：《生产组织和计划中的数学方法》(1939)，《经济资源的最优利用》(1959)，《经济最优决策》(1972，合著)，《最优规划文集》(1976)等。因在创建和发展线性规划方法以及革新、推广和发展资源最优利用理论方面所作出的杰出贡献，与美籍荷兰经济学家库恰林·库普曼斯(T.C.Koopmans，1910—1985)一起分享 1975 年度诺贝尔经济学奖。



随后，康托洛维奇继续踏实地迈进，他发现一系列涉及如何科学地组织和计划生产的问题，都属于线性规划问题。比如，怎样最充分地利用机器设备，如何最大限度地减少废料，最有效地使用燃料，怎样最合理地组织货物运输，最适当地安排农作物布局等。

北京大学出版社版权所有 禁止转载



习题 3

3.1 填空题

(1) 有 m 个产地 n 个销地的平衡运输问题模型具有__________变量，__________约束。

(2) 运输问题的检验数 σ_{ij} 与对偶变量 u_i、v_j 之间存在关系__________。

(3) 运输问题中 $m+n-1$ 个变量构成基变量的充要条件是__________。

(4) 表上作业法的基本思想和步骤与单纯形法类似，因而初始调运方案的给出就相当于找到一个__________。

(5) 若运输问题单位运价表的某一行元素分别加上一个常数 k，最优调运方案将肯定__________变化。

(6) 若调运方案中的某一空格检验数为 1，则在该空格的闭回路上调整单位运量而使总运费增加__________。

3.2 判断题

(1) 产地数与销地数相等的运输问题是产销平衡运输问题。 ( )

(2) 运输问题中用位势法求得的检验数不唯一。 ( )

(3) 用最小元素法求得的初始解比用西北角法得到的初始解在一般情况下更靠近最优解。 ( )

(4) 对于供过于求的不平衡运输问题，虚设的需求地点与各供应地之间运价为 M (M 为极大的正数)。 ( )

(5) 对于指派问题，在其效率矩阵$[c_{ij}]$中，各行或各列均减去一个常数，所求得的最优解不变。 ( )

3.3 表 3-47 为某一运输问题的单位运价表。

表 3-47 单位运价法

产地 \ 销地	B_1	B_2	B_3	供应量
A_1	65	80	80	150
A_2	70	100	90	200
需求量	50	200	100	

试问：(1)指出该运输问题的类型；(2)用最小元素法确定其初始调运方案；(3)分别用闭回路法和位势法计算空格检验数；(4)如果 A_1 的供应量降低到 100，建立该运输问题的数学模型；(5)如果 A_1 的供应量增加到 300，建立该运输问题的数学模型。

3.4 已知某运输问题的供需关系及单位运价表见表 3-48。要求：(1)用表上作业法找出最优调运方案；(2)分析使该最优方案不变时从 A_2 到 B_3 单位运价 c_{23} 的变化范围。

北京大学出版社版权所有 禁止转载

表 3-48 单位运价表

产地\销地	B_1	B_2	B_3	供应量
A_1	4	2	5	6
A_2	3	5	3	7
A_3	1	3	2	4
需求量	4	8	5	

3.5 在表 3-49 的运输问题中总需要量超过总供应量(方框中的数字是单位运费)。假定对销地 B_1、B_2和 B_3 未满足需要量的单位罚款成本是 5、3 和 2，试建立该问题的数学模型，并探讨能否将其转变为产销平衡运输问题。

表 3-49 单位运价表

产地\销地	B_1	B_2	B_3	供应量
A_1	5	1	7	10
A_2	6	4	6	80
A_3	3	2	5	15
需求量	75	20	50	

北京大学出版社版权所有 禁止转载

3.6 在表 3-50 不平衡运输问题中(方框中的数字是单位运费)，若产地 i 有单位物质未运出，就要发生存储成本。假定在产地 A_1、A_2 和 A_3 的单位存储成本是 5、4 和 3，又假定产地 A_2 的供应量必须全部运出，试建立该问题的数学模型，并探讨能否将其转变为产销平衡运输问题。

表 3-50 单位运价表

产地\销地	B_1	B_2	B_3	供应量
A_1	1	2	1	20
A_2	0	4	5	40
A_3	2	3	3	30
需求量	30	20	20	

3.7 有甲、乙、丙 3 个城市，每年分别需要煤炭 320、250、350 万吨，由 A、B 两个煤矿负责供应。已知煤矿年产量 A 为 400 万吨；B 为 450 万吨，从两煤矿至各城市煤炭运价(元/吨)见表 3-51。由于需求大于产量，经协商平均，甲城市必要时可少供 0～30 万吨，乙城市需求量须全部满足，丙城市需求量不少于 270 万吨。试求将甲、乙两矿煤炭全部分配出去，满足上述条件又使总运费为最低的调运方案。

表 3-51　单位运价表　　单位：元/吨

煤矿＼城市	甲	乙	丙
A	15	18	22
B	21	25	16

3.8 分配甲、乙、丙、丁 4 个人去完成 *A*、*B*、*C*、*D*、*E* 5 项任务，每个人完成各项任务的时间见表 3-52。由于任务数多于人数，故考虑：

(1) 任务 *E* 必须完成，其他 4 项中可任选 3 项完成；

(2) 其中有一人完成两项，其他每人完成一项。

试分别确定最优分配方案，使完成任务的总时间为最少。

表 3-52　每人完成各项任务的时间　　单位：小时

人员＼任务	*A*	*B*	*C*	*D*	*E*
甲	25	29	31	42	37
乙	39	38	26	20	33
丙	34	27	28	40	32
丁	24	42	36	23	45

北京大学出版社版权所有
禁止转载



实际操作训练

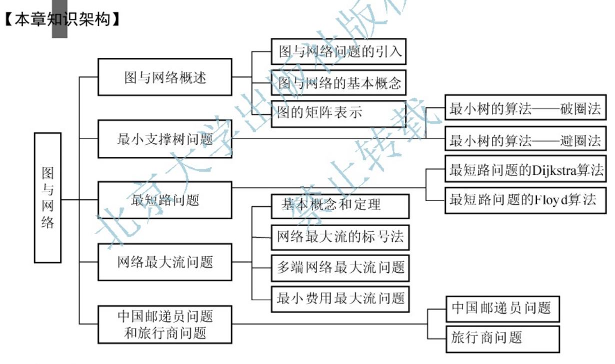
某玩具公司分别生产 3 种新型 *A*、*B* 和 *C* 玩具，每月可供量分别为 1 000 件、2 000 件、2 000 件，它们分别被送到甲、乙、丙 3 个百货商店销售。已知每月百货商店各类玩具预期销售量均为 1 500 件，由于经营方面原因，各商店销售不同玩具的盈利额不同，见表 3-53。又知丙百货商店要求至少供应 *C* 玩具 1 000 件，而拒绝进 *A* 种玩具。求满足上述条件下使总盈利额为最大的供销分配方案。

表 3-53　各商店销售玩具的盈利

玩具＼百货商店	甲	乙	丙	供应量
A	5	4	–	1 000
B	16	8	9	2 000
C	12	10	11	2 000

第4章　图与网络

【本章知识架构】



北京大学出版社版权所有 禁止转载

【本章教学目标与要求】

- 掌握图的相关概念，了解图的矩阵表示。
- 掌握求解最小树的破圈法、避圈法。
- 掌握最短路的 Dijkstra 和 Floyd 算法。
- 掌握最大流的标号算法，了解最小费用最大流问题算法。
- 了解中国邮递员问题和旅行商问题的实际背景及其求解方法。
- 针对实际情况，能够熟练识别、转换为各类典型图与网络问题，并能正确建模和求解。

导入案例

运储物流的车辆调度

运储物流公司的车辆调度(Vehicle Routing and Scheduling Problem，VRP)包括两部分内容：其一是车辆行车路线的安排；其二是出行时间表安排。一个车辆路线就是对应一系列需求点，从一个资源点出发，有序地通过它们，最后再回到出发的资源点。一个车辆的时间表就是对应一系列需求点的到达和离开的时间表，车辆必须按一定的次序在规定的时间通过这些点。由于在实际运输任务的性质和特点不同、道路条件及车辆类型不同，即使在相同收发货运点间完成同样任务时，所采用的行驶路线方案也可能不同，而车辆按不同运行路线完成同样的运输工作时，其实施效果也有所不同。因此，在满足运输任务要求前提下，如何选择最经济的运行路线，是该公司车辆调度安排的一项重要工作。

资料来源：http://define.cnki.net.

北京大学出版社版权所有 禁止转载

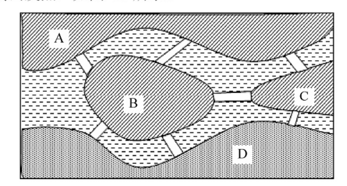
图是客观世界最直观、最简单的符号和沟通语言，用来表示所研究对象以及它们之间的某种特定的联系，在物流活动中有着广泛的应用。网络是图的一种特殊形式，诸如交通网络、水暖管道网络、局域网络等都是网络图的原型。图与网络理论在近几十年来发展很快，用来研究图与网络的一般规律以及典型问题的定量算法，这里分析几类经典的网络图问题。

4.1 图与网络概述

“图”由点和线段组成，一般用点表示这些具体事物，用连接两点的线段(直的或曲的)表示两个事物的特定的联系，就得到了描述这个“图”的几何形象。借助于图的概念、理论和方法，可对许多实际问题的建模求解。

4.1.1 图与网络问题的引入

图论的研究最早可追溯到著名的七桥问题。18 世纪欧洲的哥尼斯堡城(今俄罗斯加里宁格勒)中有一条河叫普雷格尔河，该河中有两个岛，河上有七座桥，如图 4.1 所示。当时那里的人们就考虑：能否从 A、B、C、D 中任一地方出发，每座桥走且仅走一次，最后刚好回到出发点，如图 4.1 所示。



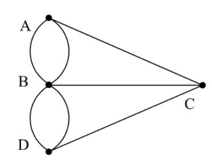


图 4.1 哥尼斯堡七桥问题

当然可以到实地去尝试解决这个问题，但该城居民的任何尝试均未成功。欧拉为了解决这个问题，采用了建立数学模型的方法。他将每一块陆地用一个点来代替，将每一座桥用连接相应两点的一条线来代替，从而得到一个有 4 个“点”、7 条“线”的“图”，问题转化为从任一点出发一笔画出七条线再回到起点。欧拉考察了一般一笔画的结构特点，给出了一笔画的一个判定法则：这个图是连通的，且每个点都与偶数条线相关联，将这个判定法则应用于七桥问题，得到了“不可能走通”的结果，不但彻底解决了这个问题，而且开创了图论研究的先河。

图与网络是运筹学中的一个经典和重要的分支，所研究的问题涉及经济管理、工业工程、交通运输、计算机科学与信息技术、通信与网络技术等诸多领域。下面再通过一些实例来了解图与网络问题。

【例 4.1】最短路问题。

一名货车司机奉命在最短的时间内将一车货物从甲地运往乙地。从甲地到乙地的公路网纵横交错，因此有多种行车路线，这名司机应选择哪条线路呢？假设货柜车的运行速度是恒定的，那么这一问题相当于需要找到一条从甲地到乙地的最短路。

【例 4.2】公路连接问题。

某一地区有若干个主要城市，现准备修建高速公路把这些城市连接起来，使得从其中任何一个城市都可以经高速公路直接或间接到达另一个城市。假定已经知道了任意两个城市之间修建高速公路的成本，那么应如何决定在哪些城市间修建高速公路，使得总成本最小？

【例 4.3】运输问题。

某种原材料有 m 个产地，现在需要将原材料从产地运往 n 个使用这些原材料的工厂。假定 m 个产地的产量和 n 家工厂的需要量已知，单位产品从任一产地到任一工厂的运费已知，那么如何安排运输方案可以使总运输成本最低？

北京大学出版社版权所有 禁止转载

【例 4.4】中国邮递员问题。

一名邮递员负责投递某个街区的邮件。如何为他(她)设计一条最短的投递路线(从邮局出发，经过投递区内每条街道至少一次，最后返回邮局)？由于这一问题是我国管梅谷教授 1960 年首先提出的，所以国际上称之为中国邮递员问题。

【例 4.5】旅行商问题。

一名推销员准备前往若干城市推销产品。如何为他(她)设计一条最短的旅行路线(从驻地出发，经过每个城市恰好一次，最后返回驻地)？这一问题的研究历史十分悠久，通常称之为旅行商问题。德国人很喜欢精确的数学，在 1978 年，波恩大学有一位数学家想知道在西德的 120 个有铁路穿过的城市要安排一个最短路程的回路，应该怎么样跑。他从铁路局找到了准确的城市间铁路的长度，整个问题变成一个有 7 140 个变量、120 个方程及 96 个不等式的线性规划问题，用电子计算机去算得到最短的回路是 6 942 千米。

【例 4.6】指派问题。

一家公司经理准备安排 n 名员工去完成 n 项任务，每人一项。由于各员工的特点不同，不同的员工去完成同一项任务时所获得的回报是不同的。如何分配工作方案可以使总回报最大？

上述问题有两个共同的特点：一是它们的目的都是从若干可能的安排或方案中寻求某种意义下的最优安排或方案；二是它们都易于用图形的形式直观地描述和表达。后面把这

种与图相关的结构称为网络(Network)，图与网络所研究的就是可用图表达的那些管理活动的优化问题。

4.1.2　图与网络的基本概念

这里所说的图与平面几何中的图不同，关心的是图中有多少个点，点与点之间有无线连接，至于连线的方式是直线还是曲线，点与点的相对位置如何，都是无关紧要的。下面介绍有关图的基本概念。

定义 4.1　图：图是点和线所组成的图形，即图是一个有序二元组(V,E)，记为$G=(V,E)$，其中$V=\{v_1,v_2,\cdots,v_p\}$是p个点的集合，$E=\{e_1,e_2,\cdots,e_q\}$是q条边的集合。V中的元素v_i称为顶点或节点，其所有元素的个数记为$|V|$，E中的元素e_k称为边。

图 4.2 所示的图：$V=\{v_1,v_2,v_3,v_4,v_5\}$，$E=\{e_1,e_2,e_3,e_4,e_5,e_6\}$，其中$e_1=(v_1,v_1)$，$e_2=(v_1,v_2)$，$e_3=(v_1,v_3)$，$e_4=(v_3,v_4)$，$e_5=(v_3,v_4)$，$e_6=(v_1,v_4)$。

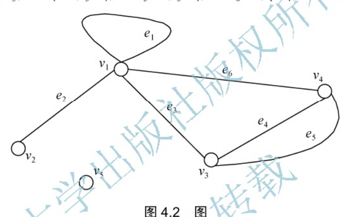


图 4.2　图

北京大学出版社版权所有
禁止转载

对于边(v_i,v_j)，称v_i、v_j为边(v_i,v_j)的端点，这时也称v_i、v_j相邻；若两条边有一个公共端点v，则称这两边相邻；若两端点之间多于一条边，称为多重边，如图 4.2 中的e_4，e_5；若一条边的两个端点相同，则称此为环(自回路)，如图 4.2 中的e_1。有时为了简便，也用顶点的编号表示顶点和边，后面这两种表示方法都会用到，不再赘述。

定义 4.2　简单图与多重图：不含环与多重边的图称为简单图，含有多重边的图称为多重图，图 4.2 就是一个多重图。

定义 4.3　无向图和有向图：图$G=(V,E)$中，若对所有的边均有$e_k=(v_i,v_j)=(v_j,v_i)$，$k=1,2,\cdots,q$，则称G为无向图，记为$G=(V,E)$。若图中边(v_i,v_j)的端点是有序的，即以v_i为来点v_j为去点，则称该图为有向图，记为$D=(V,A)$。在有向图中，把边改称为弧，A为D中弧的集合，边可以看成两条方向相反的弧。

定义 4.4　链和路：在无向图$G=(V,E)$，称一个点和边交替的序列$\{v_{i1},e_{i1},v_{i2},e_{i2},\cdots,e_{it-1},v_{it}$为连接$v_{i1}$和$v_{it}$的一条链，简记为$\{v_{i1},v_{i2},\cdots,v_{it}\}$，其中$e_{ik}=(v_{ik},v_{ik+1}),k=1,2,\cdots,t-1$。点边序列中只有重复的点而无重复边的链称为简单链；点边序列中没有重复的点和重复边者称为初等链；首尾相接的链称为圈。

图 4.3 中：$S_1=\{v_6,v_5,v_1,v_5,v_4,v_3\}$是一条连接$v_6$和$v_3$的简单链；$S_2=\{v_6,v_5,v_1,v_4,v_3\}$是一条连接$v_6$和$v_3$的初等链；$S_3=\{v_6,v_5,v_1,v_6\}$构成一个圈。

在有向图$D=(V,A)$中，称一个点和弧交替的序列$\{v_{i1},a_{i1},v_{i2},a_{i2},\cdots,a_{it-1},v_{it}\}$为连接$v_{i1}$和

v_{it} 的一条路，仍简记为 $\{v_{i1}, v_{i2}, \cdots, v_{it}\}$，若 $v_{i1} = v_{it}$，则称之为回路。

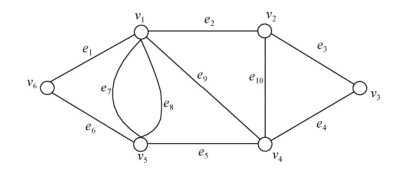


图 4.3　链

在图 4.4 中，$S_1 = \{v_6, v_5, v_1, v_5, v_4, v_3\}$ 是一条连接 v_6 和 v_3 的路；$S_2 = \{v_1, v_2, v_4, v_5, v_1\}$ 是一条回路。

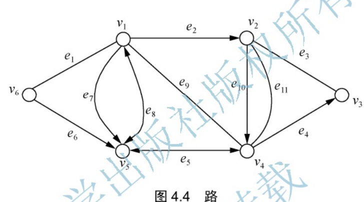


图 4.4　路

北京大学出版社版权所有
禁止转载

定义 4.5　连通图：如果一个图中任意两点间至少有一条链相连，则称此图为连通图。图 4.5(a)就是一个连通图。

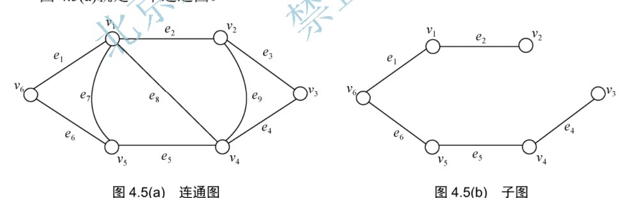


图 4.5(a)　连通图　　　　图 4.5(b)　子图

定义 4.6　子图和支撑子图：给定一个图 $G = (V, E)$，若图 $G_1 = (V_1, E_1)$ 满足 $V_1 \subset V$ 及 $E_1 \subset E$，称图 G_1 是图 G 的子图，特别地，如果 $V_1 = V$ 及 $E_1 \subset E$，则称图 G_1 是图 G 的支撑子图。

图 4.5(b)就是图 4.5(a)的子图，同时也是支撑子图。支撑子图简化了原图的结构，是最常用的子图。

定义 4.7　赋权图和网络：在实际问题中，只用图来描述所研究对象之间的关系往往是不够的，与图联系在一起，通常还有与点或边有关的某些数量指标，通常称之为“权”。权可以表示为：距离、费用、通行数量等，称含有数量指标的图为赋权图，记为 $G=(V, E, W)$。

网络一般是指一个弧上有某种所谓“流转物”流动的有向赋权图，如交通网中的车辆和管道流动的油气，记为$D=(V,A,W)$。

由于无向图也可看成为有向图，在不引起混淆的前提下，无向图的有关概念可用于有向图的描述中，这样赋权图与网络的区别并不大。

4.1.3 图的矩阵表示

用矩阵来组织图的信息使用起来非常便利，和图有关的矩阵有多种形式，这里给出其中两种。

1. 邻接矩阵

相应于图$G=(V,E)$(或D)构造一个矩阵$\boldsymbol{A}=[a_{ij}]_{n\times n}$，其中$a_{ij}=\begin{cases}1, & (v_i,v_j)\in E\\0, & 其他\end{cases}$，则称矩阵$\boldsymbol{A}$为图$G$ (或D) 的邻接矩阵。无向图的邻接矩阵是对称矩阵。

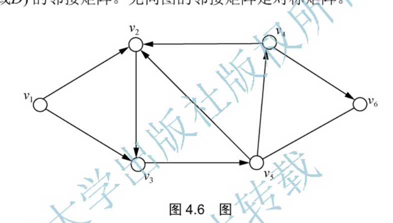


图 4.6　图

图 4.6 的邻接矩阵为

$$\boldsymbol{A}=\begin{bmatrix}0&1&1&0&0&0\\0&0&1&0&0&0\\0&0&0&0&1&0\\0&1&0&0&0&1\\0&1&0&1&0&1\\0&0&0&0&1&0\end{bmatrix}$$

给出了一个图的邻接矩阵就等于给出了图的结构信息，可以从邻接矩阵中得到图的很多重要特征，如路径问题。图 4.6 中由邻接矩阵就可算出D中任一点与其他点之间是否有路可通？若有路，走几步可以达到该点？

下面通过邻接矩阵的计算来求解$v_1\to v_5$和$v_1\to v_6$有无路可通。

先求$\boldsymbol{A}^2$，得

$$\boldsymbol{A}^2=\boldsymbol{A}\boldsymbol{A}=[a_{ij}^{(2)}]=\begin{bmatrix}0&1&1&0&1&0\\0&1&0&0&1&0\\0&1&1&1&0&1\\0&0&1&0&1&0\\0&1&1&0&1&1\\0&1&0&1&0&1\end{bmatrix}$$

北京大学出版社版权所有
禁止转载

其中：$a_{ij}^{(2)} = \sum_{i=1}^{6} a_{ik} a_{kj}$。

可以理解，若 $a_{i1}a_{1j} = 1$ 时，当且仅当 a_{i1} 和 a_{1j} 同时等于 1，所以 $a_{i1}a_{1j} = 1$ 表示从 v_i 到 v_j 有一条路，而 $a_{ij}^{(2)} = 1$ 则表示从 v_i 到 v_j 可两步到达；$a_{16}^{(2)} = 0$，表明 v_1 到 v_6 两步不能到达，继续计算 $\boldsymbol{A}^3$，得

$$\boldsymbol{A}^3 = \boldsymbol{A}^2\boldsymbol{A} = [a_{ij}^{(3)}] = \begin{bmatrix} 0 & 2 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 2 & 1 & 0 & 2 & 1 \\ 0 & 2 & 0 & 1 & 1 & 1 \\ 0 & 2 & 1 & 1 & 2 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}$$

由于 $a_{16}^{(3)} = 1$，表明从 v_1 三步可达 v_6，若要了解这条路沿途经过哪些顶点到达 v_6，只要回溯前面计算过程中的 $a_{16}^{(3)}$ 这个数是怎样计算的。因为 $a_{16}^{(3)}$ 是由 $\boldsymbol{A}^2$ 中的第一行与 $\boldsymbol{A}$ 中的第六列相应各数乘加而得，即是由 $a_{15}^{(2)} = 1$ 和 $a_{56} = 1$ 相乘而得，而 $a_{15}^{(2)} = 1$ 由 $a_{13} = 1$ 与 $a_{35} = 1$ 相乘而得，因此可知路径为 $v_1 \to v_3 \to v_5 \to v_6$。

2. 权矩阵

网络 $G = (V, E, W)$(或D)，其边 (v_i, v_j) 的权重为 w_{ij}，构造矩阵 $\boldsymbol{A} = \left[a_{ij}\right]_{n \times n}$，其中 $a_{ij} = \begin{cases} W_{ij}, & (v_i, v_j) \in E \\ \infty, & \text{其他} \end{cases}$，称矩阵 $\boldsymbol{A}$ 为网络 G(或D) 的权矩阵，其中主对角线上的元素 a_{ij} 均为零。

北京大学出版社版权所有 禁止转载

图 4.7 的权矩阵为

$$\boldsymbol{A} = \begin{bmatrix} 0 & 9 & 2 & 4 & 7 \\ 9 & 0 & 3 & 4 & \infty \\ 2 & 3 & 0 & 8 & 5 \\ 4 & 4 & 8 & 0 & 6 \\ 7 & \infty & 5 & 6 & 0 \end{bmatrix}$$

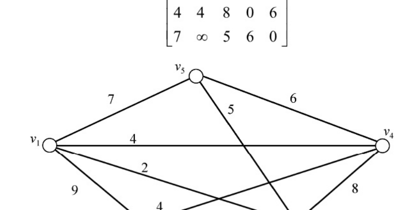


图 4.7　网络

【例 4.7】有甲、乙、丙、丁、戊、己 6 名运动员报名参加 A、B、C、D、E、F 6 个项目比赛，报名表见表 4-1，问如何安排比赛顺序，使得每个运动员不连续参加两项比赛？

表 4-1　报表

运动员＼项目	A	B	C	D	E	F
甲	√			√		√
乙	√	√		√		
丙			√		√	
丁	√				√	
戊	√	√			√	
己			√	√		√

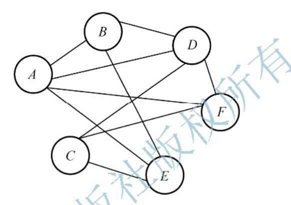


图 4.8　例 4.7 图

解：将比赛项目作为点，若两个项目有同一运动员参加，则用边连接，表示两个项目连续安排，如图 4.8 所示。然后任取一点开始安排项目，使第二点与前一点不相邻，如从 A 点开始，可以安排项目比赛顺序为 *ACBFED*。

北京大学出版社版权所有 禁止转载

4.2　最小支撑树问题

在各种各样的图中，有一类图极为简单却很实用，这就是树图。

定义 4.8　树与支撑树：一个无圈的连通图称为树；若该连通图还是一个支撑子图，则称这棵树为支撑树，也叫生成树。

【例 4.8】6 个城镇 v_1，v_2，v_3，v_4，v_5，v_6 的一个通信网图是一棵树，如图 4.9(a)所示，它是图 4.5(a)的支撑子图，也是其一棵支撑树；图 4.9(b)是一棵树，但不是图 4.5(a)的支撑树。

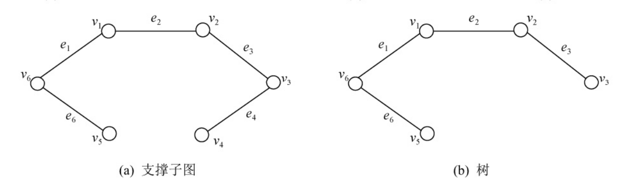


图 4.9　通信网图

容易得出树具有以下重要结论。

定理 4.1 设 $T=(V,E)$ 是一个 $|V|\geqslant 3$ 的无向图，则下列关于树的命题是等价的：

(1) T 连通且无圈；

(2) T 的任何两个顶点间均必有一条且仅有一条通路相连；

(3) T 连通且有 $n-1$ 条边，这里 $n=|V|$；

(4) T 有 $n-1$ 条边且无圈，这里 $n=|V|$；

(5) T 无圈，但在 T 中任意两个不相邻的顶点间添加一条边，就可构成一条回路；

(6) T 连通，但去掉任意一条边后就不连通，即树 T 是连通且边数最少的图。

定义 4.9 最小支撑树：设 $G=(V,E,W)$ 是一个连通的赋权图，$T=(V,S)$ 是 G 的支撑树，把 T 中所有边的权之和称作树 T 的权，记作 $w(T)$，即

$$w(T)=\sum_{e\in S}w(e)$$

G 中权最小的支撑树 T 称为 G 的最小支撑树，简称最小树。

在实际生活中，往往想知道这样一些问题的答案，例如在城镇之间建立的电话线网，如何架设使得电话线的总长最小，或总费用最小，这实际上是一个赋权图的最小树问题。下面讨论构造最小树的常用算法。

4.2.1 最小树的算法——破圈法

最常用的最小树的算法是由 Kruskal(克鲁斯卡尔)给出的破圈和避圈法，破圈法的步骤如图 4.10 所示。

(1) 在给定的赋权图上任找一个圈。

(2) 在所找的圈中去掉一条权数最大的边(若有两条或两条以上边都是权数最大的边，那么就任意去掉其中一条)。

(3) 若所剩下的图中已经不含圈，则计算停止；否则，返回(2)重新迭代计算。

直到得到一个不含圈的图为止，得到的图便是最小树。根据定理 4.1，这个过程不会一直继续下去，最后剩下 $n-1$ 条边自然结束。破圈法的思想是在保持图连通的基础上丢弃权数比较大的边，最后得到的树符合“最小”的含义，破圈法也称做丢边法。

北京大学出版社版权所有 禁止转载

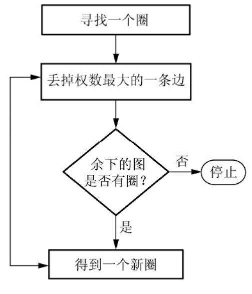


图 4.10 破圈法

【例 4.9】有 7 个单位要求煤气公司为其铺设煤气管道，经施工单位测量，7 个单位间可通管道的路线长度见表 4-2，其中的“－”表示其间无直达路线。单位 A 距煤气公司供

应网最近为 1.5 千米，现拟铺设地下管道，并经 *A* 与供应网连通，铺设费用为 25 元/米，试协助施工单位进行施工路线设计，使得费用最少，求出最小费用值。

表 4-2　路线长度　　单位：千米

i \ *j*	*A*	*B*	*C*	*D*	*E*	*F*	*G*
1：*A*	0	3	4	7	—	—	—
2：*B*	3	0	3	2	4	—	—
3：*C*	4	3	0	—	5	7	—
4：*D*	7	2	—	0	2	—	6
5：*E*	—	4	5	2	0	1	4
6：*F*	—	—	7	—	1	0	2
7：*G*	—	—	—	6	4	2	0

北京大学出版社版权所有 禁止转载

解：先建立问题的图论模型。以 7 个单位为研究对象，其间有直达路线则连一条边，在相应的边旁标以相应的路长，便构成一个网络赋权图模型如图 4.11 所示，可见其中充满了圈，从中寻求最小树便成为解题的关键。

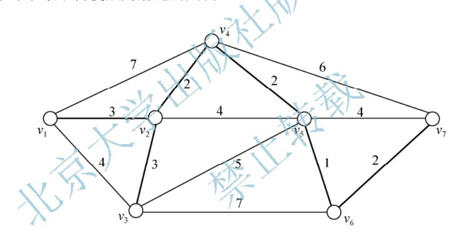


图 4.11　网络图

迭代 1：选圈 $v_1 \to v_2 \to v_3 \to v_1$，去掉权最大边 (v_1, v_3)。

迭代 2：选圈 $v_1 \to v_2 \to v_4 \to v_1$，去掉权最大边 (v_1, v_4)。

迭代 3：选圈 $v_4 \to v_5 \to v_7 \to v_4$，去掉权最大边 (v_4, v_7)。

迭代 4：选圈 $v_2 \to v_3 \to v_5 \to v_2$，去掉权最大边 (v_3, v_5)。

迭代 5：选圈 $v_2 \to v_4 \to v_5 \to v_2$，去掉权最大边 (v_2, v_5)。

迭代 6：选圈 $v_5 \to v_6 \to v_7 \to v_5$，去掉权最大边 (v_5, v_7)。

迭代 7：选圈 $v_2 \to v_3 \to v_6 \to v_5 \to v_4 \to v_2$，去掉权最大边 (v_3, v_6)。

此时图已不含圈，所得图即为最小树，其权为 13，加上 *A* 到煤气公司的 1.5 千米，总共需铺设管道长度为 14.5 千米，所需费用为(14.5×1 000×25)元=362 500 元，施工路线如图 4.12 所示。

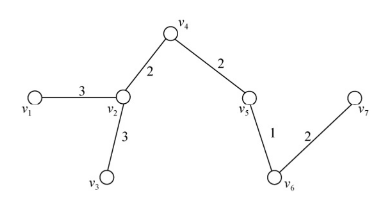


图 4.12　施工路线

4.2.2　最小树的算法——避圈法

避圈法也称为加边法，开始选一条最小权的边，以后每一步中，总从未被选中的边中选一条最小权的边，使之与已选的边不构成圈，直到所有的边都检验完，即可得最小树(每步中，若有两条或两条以上的边都是权最小的边，则从中任选一条)。其步骤如下，如图 4.13 所示。

(1) 在给定的赋权图上找一条权数最小的边。

(2) 加入这条边，构建一个新图。

(3) 若在新图中找到了 $n-1$ 条边，则计算停止；否则，在原图中找一条与新图边相邻且权数最小的边，注意该边与新图现有的边不能构成圈，返回(2)继续迭代计算。

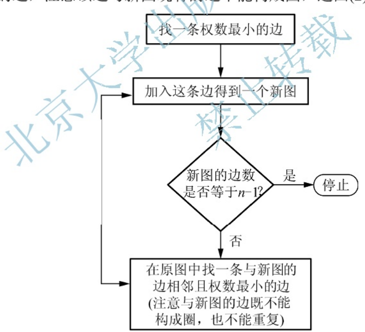


图 4.13　避圈法

北京大学出版社版权所有 禁止转载

【例 4.10】采用避圈法确定例 4.9 中的最小树。

解：为便于计算，先对图 4.11 的各边依据权数从小到大排序。

1) 迭代 1

选择边 (v_5, v_6)，构造一个新图，如图 4.14(a)所示。此时边数等于 1，小于树图要求的边数 7−1=6 (定理 4.1)，继续计算。

2) 迭代 2

选择与(v_5, v_6)相邻的边(v_2, v_5)、(v_3, v_5)、(v_4, v_5)、(v_5, v_7)、(v_3, v_6)和(v_6, v_7)中权数最小的边(v_4, v_5)或(v_6, v_7)，加入新图。既可两个边同时加入，也可任选一个，这里将这两边同时加入，以节省迭代步骤，如图 4.14(b)所示。这样新图已经有 3 个边，3 小于 6，继续迭代。

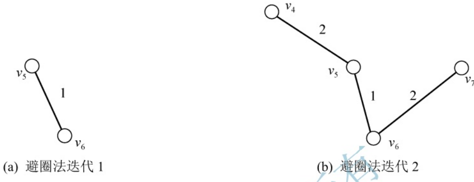


图 4.14 选择权数最小的边加入交新图

如此继续下去，直至得到 6 条边，得最小树如图 4.12 所示。

求最小树的两法各具优缺点，破圈法一般在边数比较少的情况下使用；避圈法在边数比较多的情况下使用，但它们具有共同的特征——在保持连通性的基础上去掉图中的圈。

4.3 最短路问题

北京大学出版社版权所有 禁止转载

最短路问题的一般提法是在给定网络图中找出任意两点间距离最短的一条路，可考虑有向图和无向图中的最短路。最短路问题是网络理论中应用最广泛的问题之一，许多优化问题可以使用这个模型，如管道铺设、设备更新、线路安排等。最短路问题最简单、最常用的求解方法是 Dijkstra、Floyd 算法。

定义 4.10 最短路：给定一个网络图$D=(V, A, W)$，记D中每一条弧$a_{ij}=(v_i, v_j)$上的权为$w(a_{ij})=w_{ij}$。指定D的起点v_s和终点v_t，设P是D中从v_s到v_t的一条路，则定义路P的权是P中所有弧的权之和，记为$w(P)$，即

$$w(P)=\sum_{(v_i, v_j)} w_{ij}$$

又，若P^*是D中v_s到v_t的一条路，且满足

$$w(P^*)=\min\left\{w(P)\middle| P是v_s到v_t的路\right\}$$

式中，min 表示对D的所有从v_s到 v_t的路P取最小，则称P^*为从v_s到v_t的最短路，$w(P^*)$是从v_s到v_t的最短距离。

在一个图$D=(V, A, W)$中，求从v_s到v_t的最短路或最短距离的问题称为最短路问题。

4.3.1 最短路问题的 Dijkstra 算法

Dijkstra 算法于 1959 年提出，适用于求具有非负权的有向图中的最短路，是目前公认的求无负权网络最短路问题的最好算法。

Dijkstra 算法是一种标记法，它的基本思路是从起点v_s出发，逐步向外探寻最短路，执行过程中，给每一个顶点v_j标号(λ_j, l_j)，其中λ_j是正数，它表示获得此标号的前一个点的

编号；l_j 或表示从起点 v_s 到该点 v_j 的最短路的权(称为固定标号，记为 P 标号)，或表示从起点 v_s 到该点 v_j 的最短路权的上界(称为临时标号，记为 T 标号)。该方法的每一步是去修改 T 标号，并且把某一个具有 T 标号的点改变为具有 P 标号的点，从而使图 D 中具有 P 标号的顶点数多一个，这样经过有限步，就可以求出从 v_s 到 v_t 及各点的最短路，再根据每一点的第一个数 λ_j 反向追踪找出最短路径。

Dijkstra 算法具体包括以下几个步骤。

用 P、T 分别表示某个顶点的 P 标号、T 标号，S_i 表示在第 i 步已具有 P 标号点的集合。

开始时，令 $i=0, S_0=\{v_s\}, \lambda_s=0, P(v_s)=0$，对每个 $v_j \neq v_s$，取 $T(v_j)=+\infty$，$\lambda_j=s, k=s$。

(1) 如果 $S_i=V$，算法终止。这时，对每个 $v_j \in S_i$，$l_j=P(v_j)$；否则转下一步。

(2) 设 v_k 是刚获得 P 标号的点，考察每个使 $(v_k, v_j) \in A$ 且 $v_j \notin S_i$ 的点 v_j，将 $T(v_j)$ 修改为

$$T(v_j)=\min\{T(v_j), P(v_k)+w_{kj}\}$$

如果 $T(v_j)>P(v_k)+w_{kj}$，则把 $T(v_j)$ 修改为 $P(v_k)+w_{kj}$，把 λ_j 修改为 k；否则不修改。

(3) 令 $T(v_{j_i})=\min\limits_{v_j \in \overline{S_i}}\{T(v_j)\}$，其中 $\overline{S_i}$ 是 S_i 的补集，即 $\overline{S_i}=V \setminus S_i$。如果 $T(v_{j_i})<\infty$，则把 v_{j_i} 的 T 标号变为 P 标号，即令 $P(v_{j_i})=T(v_{j_i})$，令 $S_{i+1}=S_i \cup \{v_{j_i}\}$，$k=j_i$，把 i 换成 $i+1$，返回(1)，否则终止，这时对每一个 $v_j \in S_i$，有 $l(v_j)=P(v_j)$；而对每一个 $v_j \notin S_i$，有 $l(v_j)=T(v_j)$。

【例 4.11】图 4.15 所示是某地区交通运输的示意图，试问从 v_1 出发，经哪条路线到达 v_8 才能使总行程最短？试用 Dijkstra 算法求解。

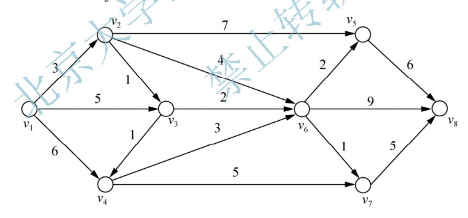


图 4.15　某地区交通运输的示意图

解：设置初始条件，令 $i=0, S_0=\{v_1\}, \lambda_1=0, P(v_1)=0$，对每个 $v_j \neq v_1$ $j=2,3,\cdots,8$，取 $T(v_j)=+\infty$，$\lambda_j=1, k=1$，即给起点 v_1 标 $(0,0)$，给其余的点标 $(1,+\infty)$，这时 v_1 为获得 P 标号的点，其余点均为 T 标号点。

1) 迭代 1($i=1$)

考察与 v_1 相邻的点 v_2, v_3和v_4 [图 4.16(a)]，因 $(v_1, v_2) \in A$，$v_2 \notin S_0$，故把 v_2 的临时标号修改为

$$T(v_2)=\min\{T(v_2), P(v_1)+w_{12}\}=\min\{+\infty, 0+3\}=3$$

这时 $\lambda_2=1$。同理，得

北京大学出版社版权所有 禁止转载

$$T(v_3)=\min\{T(v_3), P(v_1)+w_{13}\}=\min\{+\infty, 0+5\}=5, \lambda_3=1$$
$$T(v_4)=\min\{T(v_4), P(v_1)+w_{14}\}=\min\{+\infty, 0+6\}=6, \lambda_4=1$$

其余点的 T 标号不变，在所有的 T 标号中，最小的为$T(v_2)=3$，于是将v_2的标号改为P标号，即$S_1=S_0\cup\{v_2\}=\{v_1,v_2\},\lambda_2=1,P(v_2)=3,k=2,i=1$。

2) 迭代 2($i=2$)

这时v_2为刚获得 P 标号的点，考察与v_2相邻的点v_3,v_5,v_6[图 4.16(b)]，因为$(v_2,v_3)\in A$，$v_3\notin S_1$，故把v_3的临时标号修改为

$$T(v_3)=\min\{T(v_3), P(v_2)+w_{23}\}=\min\{5, 3+1\}=4$$

这时$\lambda_3=2$。同理，得

$$T(v_5)=\min\{T(v_5), P(v_2)+w_{25}\}=\min\{+\infty, 3+7\}=10, \lambda_5=2$$
$$T(v_6)=\min\{T(v_6), P(v_2)+w_{26}\}=\min\{+\infty, 3+4\}=7, \lambda_6=2$$

在所有的 T 标号中，最小的为$T(v_3)=4$，于是将v_3的标号改为 P 标号，即$S_2=S_1\cup\{v_3\}=\{v_1,v_2,v_3\},\lambda_3=2,P(v_3)=4,k=3,i=2$。

3) 迭代 3($i=3$)

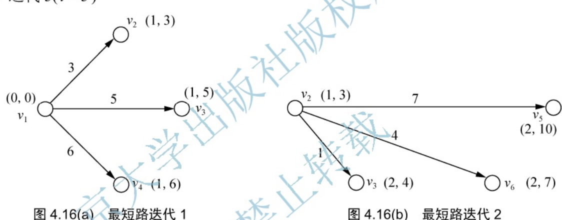


图 4.16(a) 最短路迭代 1　　图 4.16(b) 最短路迭代 2

北京大学出版社版权所有 禁止转载

这时v_3为刚获得 P 标号的点，考察与v_3相邻的点v_4和v_6[图 4.16(c)]，因为$(v_3,v_4)\in A$，$v_4\notin S_2$，故把v_4的临时标号修改为

$$T(v_4)=\min\{T(v_4), P(v_3)+w_{34}\}=\min\{6, 4+1\}=5$$

这时$\lambda_4=3$。同理，得

$$T(v_6)=\min\{T(v_6), P(v_3)+w_{36}\}=\min\{7, 4+2\}=6, \lambda_6=3$$

在所有的 T 标号中，最小的为$T(v_4)=5$，于是将v_4的标号改为 P 标号，即$S_3=S_2\cup\{v_4\}=\{v_1,v_2,v_3,v_4\},\lambda_4=3,P(v_4)=5,k=4,i=3$。

4) 迭代 4($i=4$)

这时v_4为刚获得 P 标号的点，考察与v_4相邻的点v_6和v_7[图 4.16(d)]，因$(v_4,v_6)\in A$，$v_6\notin S_3$，故把v_6的临时标号修改为

$$T(v_6)=\min\{T(v_6), P(v_4)+w_{46}\}=\min\{6, 5+3\}=6$$

这时λ_6不修改，仍等于 3。同理，得

$$T(v_7)=\min\{T(v_7), P(v_4)+w_{47}\}=\min\{+\infty, 5+5\}=10, \lambda_7=4$$

在所有的 T 标号中，最小的为$T(v_6)=6$，于是将v_6的标号改为 P 标号，即$S_4=S_3\cup\{v_6\}=\{v_1,v_2,v_3,v_4,v_6\},\lambda_6=3,P(v_6)=6,k=6,i=4$。

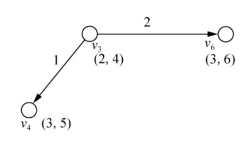


图 4.16(c) 最短路迭代 3

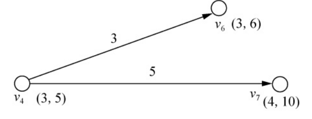


图 4.16(d) 最短路迭代 4

5) 迭代 5($i=5$)

这时 v_6 为刚获得 P 标号的点，考察与 v_6 相邻的点 v_5, v_7和v_8 [图 4.16(e)]，因 $(v_6, v_5) \in A$，$v_5 \notin S_4$，故把 v_5 的临时标号修改为

$$T(v_5) = \min\{T(v_5), P(v_6) + w_{65}\} = \min\{10, 6+2\} = 8$$

这时 $\lambda_5 = 6$。同理，得

$$T(v_7) = \min\{T(v_7), P(v_6) + w_{67}\} = \min\{10, 6+1\} = 7, \lambda_7 = 6$$

$$T(v_8) = \min\{T(v_8), P(v_6) + w_{68}\} = \min\{+\infty, 6+9\} = 15, \lambda_8 = 6$$

在所有的 T 标号中，最小的为 $T(v_7) = 7$，于是将 v_7 的标号改为 P 标号，即 $S_5 = S_4 \cup \{v_7\} = \{v_1, v_2, v_3, v_4, v_6, v_7\}, \lambda_7 = 6, P(v_7) = 7, k = 7, i = 5$。

6) 迭代 6($i=6$)

这时 v_7 为刚获得 P 标号的点，考察与 v_7 相邻的点 v_8 [图 4.16(f)]，因 $(v_7, v_8) \in A$，$v_8 \notin S_5$，故把 v_8 的临时标号修改为

$$T(v_8) = \min\{T(v_8), P(v_7) + w_{78}\} = \min\{15, 7+5\} = 12$$

这时 $\lambda_8 = 7$。在所有的 T 标号中，最小的为 $T(v_5) = 8$，于是将 v_5 的标号改为 P 标号，即 $S_6 = S_5 \cup \{v_5\} = \{v_1, v_2, v_3, v_4, v_6, v_7, v_5\}, \lambda_5 = 6, P(v_5) = 8, k = 5, i = 6$。

北京大学出版社版权所有 禁止转载

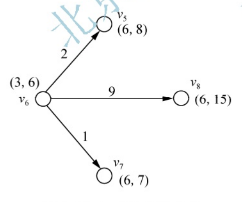


图 4.16(e) 最短路迭代 5

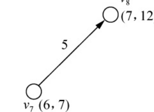


图 4.16(f) 最短路迭代 6

7) 迭代 7($i=7$)

这时 v_5 为刚获得 P 标号的点，考察与 v_5 相邻的点 v_8，因 $(v_5, v_8) \in A$，$v_8 \notin S_5$，故把 v_8 的临时标号修改为

$$T(v_8) = \min\{T(v_8), P(v_5) + w_{58}\} = \min\{12, 8+6\} = 12$$

这时 λ_8 的值不改变。在所有的 T 标号中，最小的为 $T(v_8) = 12$，于是将 v_8 的标号改为 P 标号，即 $S_7 = S_6 \cup \{v_8\} = \{v_1, v_2, v_3, v_4, v_6, v_7, v_5, v_8\}, \lambda_8 = 7, P(v_8) = 12, k = 8, i = 7$。

v_8 获得了 P 标号，迭代终止，对每一个 $v_j \in S_7$，有 $l(v_j) = P(v_j)$，如 $l(v_8) = 12$，而对

每一个 $v_j \notin S_7$，有 $l(v_j)=T(v_j)$。

至此已得到从起点 v_1 到终点 v_8 的最短距离为 $l(v_8)=12$，再根据第一个标号 λ_j 反向追踪求出最短路径为：$v_1 \to v_2 \to v_3 \to v_6 \to v_7 \to v_8$。

事实上，按照 Dijkstra 算法，也找出了起点 v_1 到各个中间点的最短路径和最短距离，例如：

$$v_1 \to v_2 \to v_3 \to v_6 \to v_5$$

就是从 v_1 到 v_5 最短路径，距离为 8。

Dijkstra 算法的思想是按距 v_s 从近到远的顺序，依次求得 v_s 到 D 的各顶点的最短路径和距离，直至 v_t (或直至 D 的所有顶点)，算法结束，为避免重复并保留每一步的计算信息。因为由起点 v_s 到 v_s 的最短路径必是由 v_s 到某个 v_k 的最短路径再加上弧 (v_k, v_t)，v_k 是 v_s 到 v_t 最短路径的点，不妨设它是 v_t 的来点，显然，$l_k = P(v_k)$ 是 v_s 到 v_k 最短路径的长度，以此类推，每个顶点的标号，一是要记录路径的来点，一是要记录最短路径的长度信息，即可获得最短路径与长度的信息。

这样，上面的迭代过程就可以简化，可采用双标号实现上面的迭代流程，且能用其求解无向图的最短路径。

(1) 给起点 v_s 的标号 $(s, 0)$，$I=\{v_s\}$，$J=V-\{v_s\}$，V 是所有顶点的集合。

(2) 基于标号集合 I，未标号集合 J，找出弧或边的集合 $A=\{(v_i, v_j) | v_i \in I, v_j \in J\}$。

(3) 如果 $A=\varnothing$，则计算结束。如果 v_t 已经标号 (k_t, l_t)，则从 v_s 到 v_t 的最短路长为 l_t，而最短路径可从 k_t 点反向追踪到起点 v_s 获得。如果 v_t 未标号，则可以判定不存在从 v_s 到 v_t 的最短路径。如果弧的集合 $A \neq \varnothing$，则转下一步。

(4) 对集合 A 中的每一条弧，计算

$$\min_{i\in I}\{l_i + w_{ij} | v_j \in J\} = l_h + w_{hk} = l_k \tag{4-1}$$

则给顶点 v_k 标号为 (h, l_k)，$I=I+\{v_k\}$，$J=J-\{v_k\}$，返回步骤(2)。

【例 4.12】 甲、乙两地之间的公路网络如图 4.17 所示。电信公司准备在甲、乙两地沿公路沿线架设一光缆线，问应如何架设可使光缆线路总长度最短？

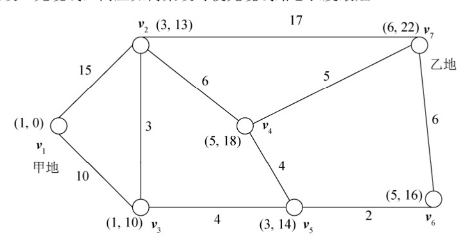


图 4.17 标号过程

解：设置初始条件，给起点 v_1 的标号 $(1, 0)$，$I=\{v_1\}$，$J=V-\{v_1\}$。

1) 迭代 1

北京大学出版社版权所有 禁止转载

考察与 v_1 相邻的点 v_2, v_3 ，得 $A=\{(v_1, v_2), (v_1, v_3)\}$，

$$\min\{l_1 + w_{12}, l_1 + w_{13}\} = \min\{0+15, 0+10\} = l_1 + w_{13} = 10$$

给顶点 v_3 标号为 $(1,10)$， $I = \{v_1, v_3\}$， $J = \{v_2, v_4, v_5, v_6, v_7\}$。

2) 迭代 2

考察与 v_1 相邻的点 v_2，与 v_3 相邻的点 v_2, v_5，得 $A=\{(v_1, v_2), (v_3, v_2), (v_3, v_5)\}$，

$$\min\{l_1 + w_{12}, l_3 + w_{32}, l_3 + w_{35}\} = \min\{0+15, 10+3, 10+4\} = l_3 + w_{32} = 13$$

给顶点 v_2 标号为 $(3,13)$， $I = \{v_1, v_3, v_2\}$， $J = \{v_4, v_5, v_6, v_7\}$。

3) 迭代 3

考察与 v_2 相邻的点 v_4, v_7，与 v_3 相邻的点 v_5，得 $A=\{(v_2, v_4), (v_2, v_7), (v_3, v_5)\}$，

$$\min\{l_2 + w_{24}, l_2 + w_{27}, l_3 + w_{35}\} = \min\{13+6, 13+17, 10+4\} = l_3 + w_{35} = 14$$

给顶点 v_5 标号为 $(3,14)$， $I = \{v_1, v_3, v_2, v_5\}$， $J = \{v_4, v_6, v_7\}$。

4) 迭代 4

考察与 v_2 相邻的点 v_4, v_7，与 v_5 相邻的点 v_4, v_6，得 $A=\{(v_2, v_4), (v_2, v_7), (v_5, v_4), (v_5, v_6)\}$，

$$\begin{aligned}&\min\{l_2 + w_{24}, l_2 + w_{27}, l_5 + w_{54}, l_5 + w_{56}\}\\&= \min\{13+6, 13+17, 14+4, 14+2\} = l_5 + w_{56} = 16\end{aligned}$$

给顶点 v_6 标号为 $(5,16)$， $I = \{v_1, v_3, v_2, v_5, v_6\}$， $J = \{v_4, v_7\}$。

5) 迭代 5

考察与 v_2 相邻的点 v_4, v_7，与 v_5 相邻的点 v_4，与 v_6 相邻的点 v_7，得 $A=\{(v_2, v_4), (v_2, v_7), (v_5, v_4), (v_6, v_7)\}$，

$$\begin{aligned}&\min\{l_2 + w_{24}, l_2 + w_{27}, l_5 + w_{54}, l_6 + w_{67}\}\\&= \min\{13+6, 13+17, 14+4, 16+6\} = l_5 + w_{54} = 18\end{aligned}$$

给顶点 v_4 标号为 $(5,18)$， $I = \{v_1, v_3, v_2, v_5, v_6, v_4\}$， $J = \{v_7\}$。

6) 迭代 6

考察与 v_2 相邻的点 v_7，与 v_6 相邻的点 v_7，与 v_4 相邻的点 v_7，得 $A=\{(v_2, v_7), (v_6, v_7), (v_4, v_7)\}$，

$$\min\{l_2 + w_{27}, l_6 + w_{67}, l_4 + w_{47}\} = \min\{13+17, 16+6, 18+5\} = l_6 + w_{67} = 22$$

给顶点 v_7 标号为 $(6,22)$， $I = \{v_1, v_3, v_2, v_5, v_6, v_4, v_7\}$， $J = \varnothing$。

至此，已找到从起点 v_1 甲地到终点 v_7 乙地的最短路径长度为 22，即架设光缆线路总长度为 22，再根据第一个标号反向追踪得出架设光缆的最短路径为： $v_1 \to v_3 \to v_5 \to v_6 \to v_7$。

4.3.2 最短路问题的 Floyd 算法

Floyd 算法是更一般的算法，该算法由 Floyd 于 1962 年提出，这是一种矩阵迭代方法。对于求任意两点间的最短路、混合图的最短路、有负权图的最短路等一般网络问题来说都比较有效。

Floyd 算法是权矩阵迭代算法，记网络的权矩阵为 $\boldsymbol{L} = [d_{ij}]_{n\times n}$， $\boldsymbol{L}$ 也是一步到达的最短距离矩阵，其中

$$d_{ij} = \begin{cases} w_{ij}, & (i, j) \in A \\ \infty, & (i, j) \notin A \end{cases}$$

北京大学出版社版权所有 禁止转载

记 $\boldsymbol{L}^m=[d_{ij}^m]_{n\times n}$ 为 m 步到达的最短距离矩阵，算法包括以下基本步骤。

(1) 令 $\boldsymbol{L}^1=\boldsymbol{L}$。

(2) 计算 $\boldsymbol{L}^k=[d_{ij}^k]_{n\times n}$，其中，$d_{ij}^k=\min\limits_{l}\{d_{il}^{k-1}+d_{lj}^{k-1}\}$。

(3) 比较矩阵 $\boldsymbol{L}^{k-1}$ 和 $\boldsymbol{L}^k$，当 $\boldsymbol{L}^{k-1}\neq\boldsymbol{L}^k$ 时，重复(2)；当 $\boldsymbol{L}^{k-1}=\boldsymbol{L}^k$ 时得到任意两点间的距离矩阵 $\boldsymbol{L}^k=[d_{ij}^k]_{n\times n}$，$d_{ij}^k$ 即为顶点 i 到 j 的最短路的距离。

若图的顶点数为 n 且 $d_{ij}\geqslant 0$，则迭代次数 k 满足

$$k-1<\frac{\lg(n-1)}{\lg 2}\leqslant k \tag{4-2}$$

注意：这里的 k 是迭代次数，不一定是网络图中某两点最短路经过的中间点的个数。

【例 4.13】用 Floyd 算法计算图 4.18 中任意两节点间的最短路。

解：根据图 4.18 写出网络图的权矩阵 $\boldsymbol{L}$。

$$\boldsymbol{L}=\begin{bmatrix}0&4&6&\infty&\infty&\infty\\4&0&2&2&1&\infty\\6&2&0&\infty&2&\infty\\\infty&2&\infty&0&1&3\\\infty&1&2&1&0&7\\\infty&\infty&\infty&3&7&0\end{bmatrix}$$

因为顶点 $n=6$，所以 $\lg 5/\lg 2\approx 2.32$，所以 $k=3$，即计算到 $\boldsymbol{L}^3$。

令：$\boldsymbol{L}^1=\boldsymbol{L}$，则

$$\boldsymbol{L}^1=\boldsymbol{L}=\begin{bmatrix}0&4&6&\infty&\infty&\infty\\4&0&2&2&1&\infty\\6&2&0&\infty&2&\infty\\\infty&2&\infty&0&1&3\\\infty&1&2&1&0&7\\\infty&\infty&\infty&3&7&0\end{bmatrix}$$

按 $d_{ij}^2=\min\limits_{l}\{d_{il}^1+d_{lj}^1\}$，$d_{ij}^3=\min\limits_{l}\{d_{il}^2+d_{lj}^2\}$ 计算得

$$\boldsymbol{L}^2=\begin{bmatrix}0&4&6&6&5&\infty\\4&0&2&2&1&5\\6&2&0&3&2&9\\6&2&3&0&1&3\\5&1&2&1&0&4\\\infty&5&9&3&4&0\end{bmatrix}\qquad \boldsymbol{L}^3=\begin{bmatrix}0&4&6&6&5&9\\4&0&2&2&1&5\\6&2&0&3&2&6\\6&2&3&0&1&3\\5&1&2&1&0&4\\9&5&6&3&4&0\end{bmatrix}$$

这里 $d_{56}^3=4$ 表示 v_5 到 v_6 的最短路长度是 4，根据其计算过程

$$\begin{aligned}d_{56}^3&=\min\left\{d_{51}^2+d_{16}^2,d_{52}^2+d_{26}^2,\cdots,d_{56}^2+d_{66}^2\right\}\\&=\min\left\{5+\infty,1+6,2+9,1+3,0+4,4+0\right\}=4=d_{56}^2+d_{66}^2\end{aligned}$$

$$\begin{aligned}d_{56}^2&=\min\left\{d_{51}^1+d_{16}^1,d_{52}^1+d_{26}^1,\cdots,d_{56}^1+d_{66}^1\right\}\\&=\min\left\{\infty+\infty,1+\infty,2+\infty,1+3,0+7,7+0\right\}=4=d_{54}^1+d_{46}^1\end{aligned}$$

北京大学出版社版权所有
禁止转载

就能确定 v_5 到 v_6 的最短路长度是 4，最短线路是 $v_5 \to v_4 \to v_6$。

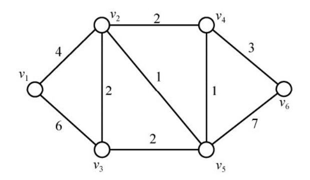


图 4.18　例 4.13 图

4.4　网络最大流问题

北京大学出版社版权所有 禁止转载

最大流问题是一类极为广泛的问题，如公路系统中有车辆流量问题，供电系统中有电流量问题，通信网络有信息流问题等。最大流问题是在单位时间内安排一个运送方案，将发点的物质沿着弧的方向运送到收点，使总运输量最大。20 世纪 50 年代，Ford(福特)、Fulkerson(富克逊)建立的“网络流理论”，是网络图应用的重要组成部分。

4.4.1　基本概念和定理

定义 4.11 容量网络和流：对网络上每一条弧都给出一个最大的通过能力，称为该弧的容量(Capacity)，记为 $c(v_i, v_j)$，简记为 c_{ij}，并称该网络为容量网络。容量网络规定一个发点 Send(源点) s，一个收点 Take(汇点) t。流(Flow)指网络上各条弧的负载量，记为 $f(v_i, v_j)$，简记 f_{ij}。

【例 4.14】 图 4.19 是连接某产品产地 v_s 和销地 v_t 的交通图。弧 (v_i, v_j) 表示从 v_i 到 v_j 的运输线，弧旁括号内第一个数字表示这条运输线的最大通过能力 c_{ij}，括号内第二个数字表示该弧上的实际流 f_{ij}。现要求制定一个运输方案，使从 v_s 运到 v_t 的产品数量最多。

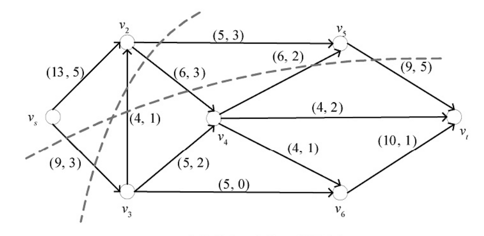


图 4.19　连接某产品产地和销地交通图

定义 4.12 可行流：对于给定的容量网络 $D=(V, A, C)$ 和给定的流 $f=\{f(v_i, v_j)\}$，若 f

满足下列条件：

(1) 容量限制条件，对每一条弧(v_i, v_j)，有$0 \leqslant f_{ij} \leqslant c_{ij}$；

(2) 流量平衡条件，对于中间点，流出量=流入量，即对于每一个$i\,(i \neq s,t)$，有

$$\sum_{(v_i, v_j)\in A} f_{ij} - \sum_{(v_j, v_i)\in A} f_{ji} = 0$$

对于发点v_s，有
$$\sum_{(v_s, v_j)\in A} f_{sj} = V(f)$$

对于收点v_t，有
$$\sum_{(v_j, v_t)\in A} f_{jt} = V(f)$$

则称$f = \{f_{ij}\}$为一个可行流，$v(f)$称为这个可行流的流量。

注意：这里所说的发点v_s是指只有从v_s发出去的弧，而没有指向v_s的弧；收点v_t只有弧指向v_t，而没有从v_t发出去的弧。

可行流总是存在的，例如，令所有弧上的流$f_{ij} = 0$，就得到一个可行流(称为零流)，其流量$v(f) = 0$。图 4.19 中，每条弧上括号内的数字给出的就是一个可行流$f = \{f_{ij}\}$，它满足定义中的条件(1)和(2)，其流量$v(f) = 5 + 3 = 8$。

所谓网络最大流问题就是求一个流$f = \{f_{ij}\}$，使得总流量$v(f)$达到最大，并且满足定义 4.12 中的条件(1)和(2)，即

$$\max v(f)$$
$$\begin{cases} \sum\limits_{(v_i, v_j)\in A} f_{ij} - \sum\limits_{(v_j, v_i)\in A} f_{ji} = 0 \\ \sum\limits_{(v_s, v_j)\in A} f_{sj} = v(f) \\ \sum\limits_{(v_j, v_t)\in A} f_{jt} = v(f) \\ 0 \leqslant f_{ij} \leqslant c_{ij}\ (i, j \neq s, t) \end{cases} \tag{4-3}$$

网络最大流问题是一个特殊的线性规划问题，可以利用线性规划方法去求解。这一节将给出一种利用图的特点来解决这个问题的方法，较线性规划的一般方法要简便和直观得多。

【例 4.15】写出图 4.19 所示的网络最大流问题的线性规划模型。

解：设该网络的可行流为$f = \{f_{ij}\}$，则其线性规划模型为

$$\max v(f)$$
$$\begin{cases} f_{s2} + f_{s3} = v(f) \\ f_{s2} + f_{32} - f_{25} - f_{24} = 0 \\ f_{s3} - f_{32} - f_{34} - f_{36} = 0 \\ f_{24} + f_{34} - f_{45} - f_{46} - f_{4t} = 0 \\ f_{25} + f_{45} - f_{5t} = 0 \\ f_{36} + f_{46} - f_{6t} = 0 \\ f_{5t} + f_{4t} + f_{6t} = v(f) \\ 0 \leqslant f_{ij} \leqslant c_{ij} \end{cases}$$

北京大学出版社版权所有 禁止转载

定义 4.13　饱和弧和零流弧：在网络 $D=(V,A,C)$ 中，若给定一个可行流 $f=\{f_{ij}\}$，将网络中 $f_{ij}=c_{ij}$ 的弧称为饱和弧，使 $0\leqslant f_{ij}<c_{ij}$ 的弧称为非饱和弧。将 $f_{ij}=0$ 的弧称为零流弧，将 $0<f_{ij}\leqslant c_{ij}$ 的弧称为非零流弧。

图 4.19 中的弧都是非饱和弧，而弧 (v_3,v_6) 为零流弧。

若 μ 是网络中连接发点 v_s 和收点 v_t 的一条路，定义路的方向是从 v_s 到 v_t，则路上的弧被分为两类。

一类是：弧的方向与路的方向一致，称此类弧为前向弧，所有前向弧的集合记为 μ^+。

一类是：弧的方向与路的方向不一致，称此类弧为后向弧，所有后向弧的集合记为 μ^-。

图 4.19 中，$\mu=\{v_s,(v_s,v_2),v_2,(v_3,v_2),v_3,(v_3,v_6),v_6,(v_6,v_t),v_t\}$ 是一条从 v_s 到 v_t 的路，则

$$\mu^+=\{\,(v_s,v_2),\ (v_3,v_6),\ (v_6,v_t)\},\quad \mu^-=\{\,(v_3,v_2)\}$$

定义 4.14　增广路：设 $f=\{f_{ij}\}$ 是网络 $D=(V,A,C)$ 上的一个可行流，μ 是从 v_s 到 v_t 的一条路，若 μ 满足下列条件：

(1) 在弧 $(v_i,v_j)\in\mu^+$ 上，其流量 $f_{ij}<c_{ij}$，即 μ^+ 中的每一条弧都是非饱和弧；

(2) 在弧 $(v_i,v_j)\in\mu^-$ 上，其流量 $f_{ij}>0$，即 μ^- 中的每一条弧都是非零流弧。则称 μ 是关于 f 的一条增广路。

如前面所说的路 $\mu=\{v_s,(v_s,v_2),v_2,(v_3,v_2),v_3,(v_3,v_6),v_6,(v_6,v_t),v_t\}$ 就是一条增广路，因为其中 μ^+ 上的弧均非饱和；而 μ^- 上的弧为非零流弧，显然这样的增广路不止一条。

定义 4.15　截集：给定网络 $D=(V,A,C)$，若点集 V 被分割成两个非空集合 V_1 和 V_2，使得 $V=V_1\cup V_2, V_1\cap V_2=\varnothing$（空集），且 $v_s\in V_1, v_t\in V_2$，则把发点在 V_1 收点在 V_2 的弧的集合称为分离 v_s 到 v_t 的一个截集，记为 (V_1,V_2)。

在图 4.19 中，设 $V_1=\{v_s,v_2,v_5\}$，$V_2=\{v_3,v_4,v_6,v_t\}$，则截集为

$$(V_1,V_2)=\{(v_s,v_3),(v_2,v_4),(v_5,v_t)\}$$

参看图 4.19 中的虚线穿过的弧，而弧 (v_3,v_2) 和 (v_4,v_5) 不是该集中的弧，因为这两条弧的起点在 V_2 中，与定义 4.15 不符。显然，一个网络的截集是很多的，在图 4.19 中还可取 $V_1'=\{v_s,v_2\}$，$V_2'=\{v_3,v_4,v_5,v_6,v_t\}$，则截集为

$$(V_1',V_2')=\{(v_s,v_3),(v_2,v_4),(v_2,v_5)\}$$

另外，若把网络 $D=(V,A,C)$ 中某截集的弧从该网络中去掉，则从 v_s 到 v_t 便不存在路，所以直观上说，截集是从 v_s 到 v_t 的必经之弧，这也是截集含义的由来。

定义 4.16　截量：在网络 $D=(V,A,C)$ 中，给定一个截集 (V_1,V_2)，则把该截集中所有弧的容量之和，称为这个截集的容量，简称为截量，记为 $c(V_1,V_2)$，即

$$c(V_1,V_2)=\sum_{(v_i,v_j)\in(V_1,V_2)}c_{ij}$$

例如在上面所举的两个截集中，有

$$c(V_1,V_2)=c_{s3}+c_{24}+c_{5t}=9+6+9=24$$

而

$$c(V_1',V_2')=c_{s3}+c_{24}+c_{25}=9+6+5=20$$

显然，截集不同截量也不同。根据截集的定义，截集的个数是有限的，故其中必有一

北京大学出版社版权所有 禁止转载

个截集的容量是最小的，称为最小截集，也就是通常所说的“瓶颈”。

关于网络最大流，有下面的重要定理。

定理 4.2 在网络 $D=(V,A,C)$ 中，可行流 $f^*=\left\{f_{ij}^*\right\}$ 是最大流的充要条件是 D 中不存在关于 f^* 的增广路。

定理 4.3 (最大流-最小截集定理)对于任意给定的网络 $D=(V,A,C)$，从发点 v_s 到收点 v_t 的最大流的流量必等于分割 v_s 到 v_t 的最小截集 (V_1^*,V_2^*) 的容量，即

$$v(f^*)=c(V_1^*,V_2^*)$$

由定理 4.2 可知，若给定一个可行流 $f=\left\{f_{ij}\right\}$，只要判断网络 D 有无关于 f 的增广路，就可判断该可行流是否为最大流。如果有增广路，则将其调整一个流量，得到新的可行流；如果没有增广路，则得到最大流。在实际计算时，用给顶点标号的方法来定义的 V_1^*，在标号过程中，有标号的顶点表示是 V_1^* 中的点，没有标号的点表示不是 V_1^* 中的点。一旦有了标号，就表明找到一条从 v_s 到 v_t 的增广路；如果标号过程无法进行下去，而 v_t 尚未标号，则说明不存在从 v_s 到 v_t 的增广路，于是得到最大流。这时将已标号的点(至少有一个点 v_s)放在集合 V_1^* 中，将未标号点(至少有一个点 v_t)放在集合 V_2^* 中，就得到一个最小截集 (V_1^*,V_2^*)。

4.4.2　网络最大流的标号法

最大流的标号法是 Ford 于 1956 年提出的，从一个可行流出发 $f=\left\{f_{ij}\right\}$(若网络中没有给定 f，则可以设 f 是零流)，经过标号过程与调整过程反复迭代，最终得到最大流。

1. 标号过程

在这个过程中，网络中的点或者是标号点(又分为已检查和未检查两种)，或者是未标号点，每个标号点的标号包含两部分：第一个标号表明它来源于哪一点，以便找出增广路；第二个标号为了确定增广路的调整量 θ。

标号过程开始，总先给 v_s 标上(0，+∞)，这时 v_s 是标号而未检查的点，其余都是未标号点。一般地，取一个标号而未检查的点 v_i，对一切未标号点 v_j，有以下关系。

(1) 在弧上 (v_i,v_j)，若 $f_{ij}<c_{ij}$，则给 v_j 标号 $(i^+,l(v_j))$，这里 i^+ 表示是前向弧上的来点编号，$l(v_j)=\min\left\{l(v_i),c_{ij}-f_{ij}\right\}$，这时点 v_j 成为标号而未检查的点。

(2) 在弧上 (v_j,v_i)，若 $f_{ji}>0$，则给 v_j 标号 $(i^-,l(v_j))$，这里 i^- 表示是后向弧上的来点编号，$l(v_j)=\min\left\{l(v_i),f_{ji}\right\}$，这时点 v_j 成为标号而未检查的点。

于是 v_i 成为标号而已检查过的点，重复上述步骤，一旦 v_t 被标上号，表明得到一条从 v_s 到 v_t 的增广路 μ，转入调整过程。

若所有标号都是已检查过，而标号过程进行不下去时，则算法结束，这时的可行流就是最大流。

北京大学出版社版权所有 禁止转载

2. 调整过程

首先按 v_t 及其他点的第一个标号，利用“反向追踪”的办法，找出增广路 μ 。例如，若 v_t 的第一个标号为 i^+ (或 i^-)，则弧 (v_i, v_t) (或相应地 (v_t, v_i))是 μ 上的弧，再检查 v_i 的第一个标号，一直进行下去，直到 v_s 为止。令调整量 $\theta = l(v_t)$ ，即 v_t 的第二个标号，调整网络中弧的流量，即

$$f'_{ij} = \begin{cases} f_{ij} + \theta & (v_i, v_j) \in \mu^+ \\ f_{ji} - \theta & (v_j, v_i) \in \mu^- \\ f_{ij} & (v_i, v_j) \notin \mu \end{cases} \tag{4-4}$$

得到新的可行流。去掉所有的标号，对新的可行流 $f' = \{f'_{ij}\}$ ，重新进行标号过程。

【例 4.16】用标号法求图 4.19 所示的网络最大流。

解：

1) 对图 4.19 中各顶点进行标号

首先给 v_s 标(0，+∞)，即 $l(v_s) = +\infty$

检查 v_s ：在弧 (v_s, v_2) 上， $5[f_{s2}] < 13[c_{s2}]$ ，

$$l(v_2) = \min\{l(v_s), c_{s2} - f_{s2}\} = \min\{+\infty, 13-5\} = 8$$

所以给 v_2 标号 $(s^+, 8)$ ，其中第一个标号表示弧 (v_s, v_2) 的端点是 v_s ，第二个标号表示该弧的可调整量是 8 个单位；

在弧 (v_s, v_3) 上，因为 $3[f_{s3}] < 9[c_{s3}]$ ，

$$l(v_3) = \min\{l(v_s), c_{s3} - f_{s3}\} = \min\{+\infty, 9-3\} = 6$$

所以给 v_3 标号 $(s^+, 6)$ 。

检查 v_2 ：在弧 (v_2, v_5) 上， $3[f_{25}] < 5[c_{25}]$ ，

$$l(v_5) = \min\{l(v_2), c_{25} - f_{25}\} = \min\{8, 5-3\} = 2$$

所以给 v_5 标号 $(2^+, 2)$ ；

在弧 (v_2, v_4) 上， $3[f_{24}] < 6[c_{24}]$ ，

$$l(v_4) = \min\{l(v_2), c_{24} - f_{24}\} = \min\{8, 6-3\} = 3$$

所以给 v_4 标号 $(2^+, 3)$ ；

在弧 (v_3, v_2) 上， $1[f_{32}] > 0$ ，

$$l(v_3) = \min\{l(v_2), f_{32}\} = \min\{8, 1\} = 1$$

所以给 v_3 标号 $(2^-, 1)$ 。

注意，前面已给 v_3 标过号 $(s^+, 6)$ ，现在又给 v_3 标号 $(2^-, 1)$ ，它们分别隶属于两条不同的路线，不存在修改标号的问题(与最短路不同)。因为要尽快找出一条从 v_s 到 v_t 的增广路，即尽快使终点 v_t 获得标号，所以不必在中途过多停留。也就是说在对已标号点 v_i 进行检查时，每次只检查一个相邻点 v_j (不论前向弧或后向弧均可)，再给标号即可，而不必检查所有与 v_i 相邻的点。事实上，其余的相邻点也不会漏掉，因为以后还要通过检查这些点来找出新的增广路。

检查 v_5 ：在弧 (v_5, v_t) 上， $5[f_{5t}] < 9[c_{5t}]$

$$l(v_t) = \min\{l(v_5), c_{5t} - f_{5t}\} = \min\{2, 9-5\} = 2$$

北京大学出版社版权所有 禁止转载

所以给 v_t 标号 $(5^+,2)$ 。

至此，终点 v_t 已获得标号，于是找出一条 v_s 到 v_t 的增广路。再由第一个标号反向追踪找出路线，即 $\mu_1=\{v_s,v_2,v_5,v_t\}$ ，如图 4.20(a)所示。

2) 进行调整

这时的调整量 $\theta=l(v_t)=2$ ，按照公式(4-4)调整。由于 μ_1 上各弧均为前向弧，故

$$f'_{s2}=f_{s2}+\theta=5+2=7;f'_{25}=f_{25}+\theta=3+2=5;f'_{5t}=f_{5t}+\theta=5+2=7$$

其余的 f_{ij} 不变，这样就得到新的可行流，如图 4.21 所示。

对这个新的可行流再进入标号过程，寻找增广路。开始给 v_s 标(0，+∞)，即 $l(v_s)=+\infty$

检查 v_s ：在弧 (v_s,v_2) 上， $7<13$ ，

$$l(v_2)=\min\{l(v_s),c_{s2}-f_{s2}\}=\min\{+\infty,13-7\}=6$$

所以给 v_2 标号 $(s^+,6)$ 。

检查 v_2 ：在弧 (v_3,v_2) 上， $1>0$ ，

$$l(v_3)=\min\{l(v_2),f_{32}\}=\min\{6,1\}=1$$

所以给 v_3 标号 $(2^-,1)$ 。

检查 v_3 ：在弧 (v_3,v_6) 上， $0<5$ ，

$$l(v_6)=\min\{l(v_3),c_{36}-f_{36}\}=\min\{1,5-0\}=1$$

所以给 v_6 标号 $(3^+,1)$ 。

检查 v_6 ：在弧 (v_6,v_t) 上， $1<10$ ，

$$l(v_t)=\min\{l(v_6),c_{6t}-f_{6t}\}=\min\{1,10-1\}=1$$

所以给 v_t 标号 $(6^+,1)$ 。

至此，终点 v_t 已获得标号，于是找出一条 v_s 到 v_t 的增广路。再由第一个标号反向追踪找出路线，即 $\mu_2=\{v_s,v_2,v_3,v_6,v_t\}$ ，如图 4.20(b)所示。

这时的调整量 $\theta=l(v_t)=1$ ，按照公式(4-4)调整。由于 μ_2 上既有前向弧又有后向弧，逐条调整

$$f'_{s2}=f_{s2}+\theta=7+1=8;f'_{32}=f_{32}-\theta=1-1=0$$
$$f'_{36}=f_{36}+\theta=0+1=1;f'_{6t}=f_{6t}+\theta=1+1=2$$

其余的 f_{ij} 不变，这样就得到了一个新的可行流，重新标号调整。

同理，开始给 v_s 标号，得到一条增广路 $\mu_3=\{v_s,v_2,v_4,v_t\}$ ，进行调整，如图 4.20(c)所示，再重新进行标号寻找新的增广路。

开始给 v_s 标号，得到一条增广路 $\mu_4=\{v_s,v_2,v_4,v_6,v_t\}$ ，进行调整，见图 4.20(d)，再重新进行标号寻找新的增广路。

开始给 v_s 标号，得到一条增广路 $\mu_5=\{v_s,v_3,v_4,v_6,v_t\}$ ，进行调整，如图 4.20(e)所示，再重新进行标号寻找新的增广路。

开始给 v_s 标号，得到一条增广路 $\mu_6=\{v_s,v_3,v_6,v_t\}$ ，进行调整，如图 4.20(f)所示，再重新进行标号寻找新的增广路。

北京大学出版社版权所有
禁止转载

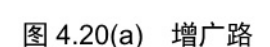


图 4.20(a)　增广路

图 4.20(b)　增广路

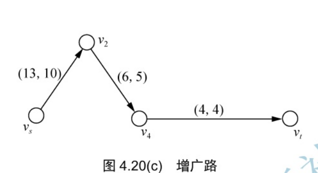


图 4.20(c)　增广路

图 4.20(d)　增广路

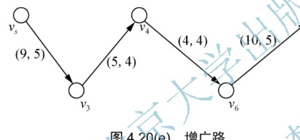


图 4.20(e)　增广路

图 4.20(f)　增广路

北京大学出版社版权所有
禁止转载

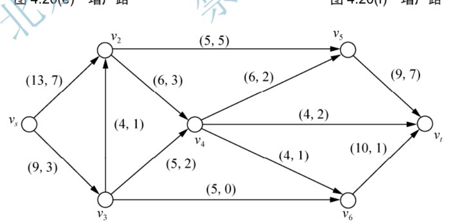


图 4.21　新的可行流

开始给 v_s 标(0，+∞)，检查 v_3，这时弧 (v_s, v_3) 已饱和，标号无法进行；检查 v_2，还可以标号 $(s^+, 2)$；再检查 v_4、v_5，标号也无法进行，最后得到最大流 $f^* = \{f_{ij}^*\}$，如图 4.22 所示，最大流量为

$$v(f^*) = f_{s2}^* + f_{s3}^* = f_{4t}^* + f_{5t}^* + f_{6t}^* = 20$$

与此同时，也可找到最小截集 (V_1^*, V_2^*)，其中 V_1^* 为最后一轮已标号点的集合，V_2^* 为未

标号点的集合，即

$$V_1^* = \{v_s, v_2\}; \qquad V_2^* = \{v_3, v_4, v_5, v_6, v_t\}$$
$$(V_1^*, V_2^*) = \{(v_s, v_3), (v_2, v_4), (v_2, v_5)\}$$

最小截量为 $c(V_1^*, V_2^*) = c_{s3} + c_{24} + c_{25} = 9 + 6 + 5 = 20$，如图 4.22 中的虚线。

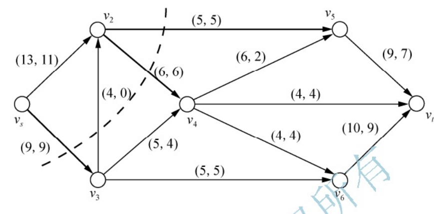


图 4.22 网络

由上述过程可见，用标号法找增广路求最大流的同时也得到一个最小截集。最小截集容量的大小影响总运输量的提高，因此，为提高总运输量，必须首先考虑增大最小截集中各弧的容量，提高它们的通过能力。反之，一旦最小截集中弧的通过能力降低，必然会使得运输量减少。

4.4.3 多端网络最大流问题

前面讨论都是对一个发点、一个收点的网络最大流问题，对于多个发点和收点的情形，可以采取虚设一个总发点 v_s 和总收点 v_t，从总发点 v_s 到各发点 v_{si} 均以弧相连，并且令这些弧的容量均为 ∞ 或某一具体值(根据情况而定，以不改变网络的通行能力为基准)；同样，从各个收点 v_{jt} 到总收点 v_t 也以弧相连，也令这些弧的容量为 ∞ 或某一具体值。这样，原来的发点 v_{si} 与收点 v_{jt} 都变成了转运点，原来的问题就转变成一个发点、一个收点的网络图，采用标号法进行求解即可。

例如图 4.23 所示是两个发点两个收点的网络，可以转换成一个发点、一个收点的网络。

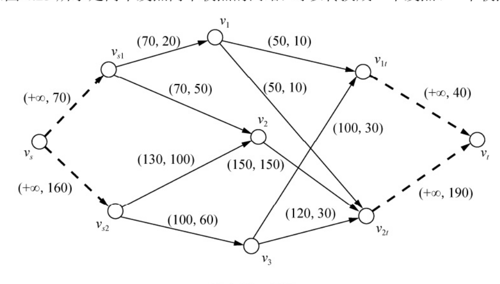


图 4.23 网络

北京大学出版社版权所有 禁止转载

注意添加虚拟弧的容量和流量的限制，一般容量要尽量选大一些，流量必须满足平衡条件。

4.4.4　最小费用最大流问题

网络最大流问题仅考虑两点间容量限制条件下的最大调配方案，不考虑各弧的费用；运输问题研究产销平衡时费用最小的调配方案，不考虑产销点之间的容量限制。最小费用最大流问题既要考虑各弧费用，又要考虑各弧的容量，运输问题、最大流问题、最短路问题(距离代表单位流量的费用)都是其特例。

1. 费用的有关概念

设 $D=(V, A, C)$ 为一给定网络，对任意 $(v_i, v_j)\in A$ 除了给出容量 c_{ij} 外，还给出了这段弧的单位流量的费用 b_{ij}。

定义 4.17　可行流的费用：对 D 上的一个可行流 $f=\{f_{ij}\}$，称 $B(f)=\sum\limits_{(v_i, v_j)\in A} b_{ij}f_{ij}$ 为可行流 f 的费用。

所谓最小费用最大流问题就是当最大流不唯一时，在这些最大流中求一个流，使得该流的总费用最小。下面介绍解决最小费用最大流问题的方法，其基本思想是在寻求最大流的算法过程中考虑费用最小的流。

定义 4.18　增广路的费用：μ 是一条关于可行流 f 的增广路，若以 θ 调整 f，得到新的可行流 f'，将

$$\begin{aligned}B(f')-B(f)&=\sum_{(v_i, v_j)\in A} b_{ij}f'_{ij}-\sum_{(v_i, v_j)\in A} b_{ij}f_{ij}\\&=\sum_{(v_i, v_j)\in \mu} b_{ij}f'_{ij}-\sum_{(v_i, v_j)\in \mu} b_{ij}f_{ij}\\&=\left[\sum_{(v_i, v_j)\in \mu^+} b_{ij}f'_{ij}-\sum_{(v_i, v_j)\in \mu^+} b_{ij}f_{ij}\right]+\left[\sum_{(v_i, v_j)\in \mu^-} b_{ij}f'_{ij}-\sum_{(v_i, v_j)\in \mu^-} b_{ij}f_{ij}\right]\\&=\theta\left[\sum_{(v_i, v_j)\in \mu^+} b_{ij}\right]-\theta\left[\sum_{(v_i, v_j)\in \mu^-} b_{ij}\right]=\theta\left[\sum_{(v_i, v_j)\in \mu^+} b_{ij}-\sum_{(v_i, v_j)\in \mu^-} b_{ij}\right]\end{aligned}$$

称为增广路 μ 的费用。

2. 最小费用最大流的算法

若 f 是流量为 $v(f)$ 的所有可行流中费用最小者，而 μ 是关于 f 的费用最小的增广路，那么沿 μ 去调整，得到的可行流 f'，就是流量为 $v(f')$ 的所有可行流中费用最小的流。因此最小费用最大流需先找一个最小费用流，然后找出最小费用增广路并进行调整，一直进行下去，直到找不出增广路为止，这时的可行流即为最小费用最大流。

如何寻找最小费用增广路呢？构造一个有向赋权图 $W(f)$，其顶点为原网络 D 的顶点，把 D 中的每条弧 (v_i, v_j) 变为两条相反方向的弧 (v_i, v_j)、(v_j, v_i)，并规定 $W(f)$ 中弧的权 w_{ij} 和 w_{ji} 分别为

$$w_{ij}=\begin{cases}b_{ij}, & f_{ij}<c_{ij}\\+\infty, & f_{ij}=c_{ij}\end{cases}\qquad w_{ji}=\begin{cases}-b_{ij}, & f_{ij}>0\\+\infty, & f_{ij}=0\end{cases}$$

北京大学出版社版权所有 禁止转载

长度为$+\infty$的弧在找增广路可以略去，于是求最小费用增广路等价于在$W(f)$中求从v_s到v_t的最短路。

由以上讨论可得出求最小费用最大流的算法。

(1) 第 1 步取零流为初始最小费用可行流，记为$f^{(0)}$。

(2) 若第k步得到最小费用流$f^{(k)}$，构造一个有向赋权图$W(f^{(k)})$，在$W(f^{(k)})$中寻求从v_s到v_t的最短路。若不存在最短路，则$f^{(k)}$即为网络D的最小费用最大流；若存在最短路，则在原网络D中得到了相应的最小费用增广路μ，对$f^{(k)}$做调整，调整量为

$$\theta=\min\left\{\min_{(v_i,v_j)\in\mu^+}\left\{c_{ij}-f_{ij}^{(k)}\right\},\min_{(v_i,v_j)\in\mu^-}\left\{f_{ij}^{(k)}\right\}\right\}$$

调整$f^{(k)}$的各弧，得到新的可行流$f^{(k+1)}=\left\{f_{ij}^{(k+1)}\right\}$，重新计算。

【例 4.17】 求如图 4.24 网络的最小费用最大流，弧旁的数字为(b_{ij},c_{ij})。

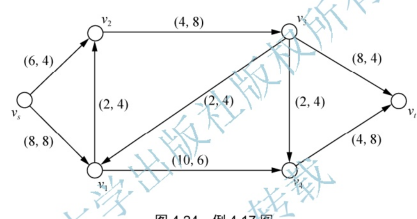


图 4.24 例 4.17 图

北京大学出版社版权所有 禁止转载

解：(1) 取$f^{(0)}$为初始可行流。

(2) 构造$W(f^{(0)})$，如图 4.25(a)所示，最短路为$\mu=\{v_s,v_1,v_3,v_4,v_t\}$，在$\mu$上进行调整，$\theta=\min\{4,8,4,8\}=4$，得到新的可行流$f^{(1)}$，如图 4.25(b)所示，弧旁括号中的数字为$(c_{ij},f_{ij})$。

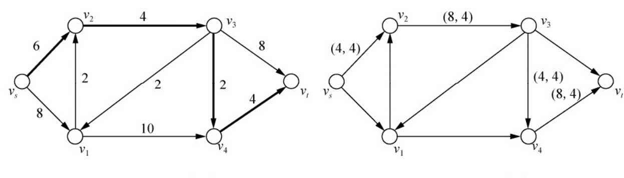


图 4.25(a) 最短路　　图 4.25(b) 可行流

(3) 构造相应的赋权有向图$W(f^{(1)})$，如图 4.25(c)所示，最短路$\mu=\{v_s,v_1,v_2,v_3,v_t\}$，在$\mu$上进行调整，得到新的可行流$f^{(2)}$，如图 4.25(d)所示。

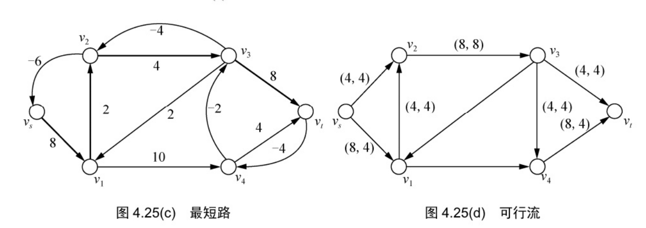


图 4.25(c) 最短路　　图 4.25(d) 可行流

(4) 构造相应的赋权有向图 $W(f^{(2)})$，如图 4.25(e)所示，最短路 $\mu=\{v_s, v_1, v_4, v_t\}$，在 μ 上进行调整，得到新的可行流 $f^{(3)}$，如图 4.25(f)所示。

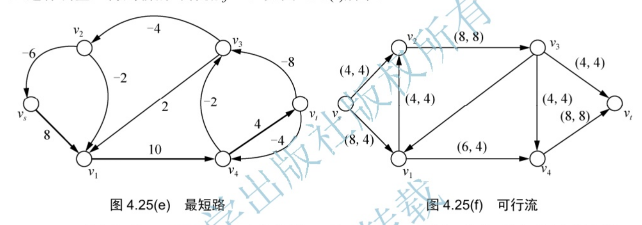


图 4.25(e) 最短路　　图 4.25(f) 可行流

(5) 可以看到 $W(f^{(3)})$ 中已不存在从 v_s 到 v_t 的路，所以 $f^{(3)}$ 为最小费用最大流，$v(f^{(3)})=12, B(f^{(3)})=240$。

北京大学出版社版权所有 禁止转载

4.5 中国邮递员问题和旅行商问题

这里简要介绍两个著名的图论问题：一是中国邮递员问题；二是旅行商问题。

4.5.1 中国邮递员问题

一名邮递员带着要分发的邮件从邮局出发，经过要分发的每个街道，送完邮件后又返回邮局，如果他必须至少一次走过他管辖范围内的每一条街道，如何选择投递路线，使邮递员走尽可能少的路程。该问题由我国数学家管梅谷在 1962 年首次提出，在国际上称为中国邮递员问题。

1. 基本概念和定理

中国邮递员问题用图论的术语来描述，就是在一个连通的赋权图 $G(V, E)$ 中寻找一条回路，使该回路包含 G 中的每条边至少一次，且该回路的权数最小，也就是说要从包含 G 的每条边的回路中找一条权数最小的回路。

定义 4.19 欧拉回路：设 $G(V, E)$ 为一个图，若存在一条回路，使它经过图中每条边且只经过一次又回到起始点，就称这种回路为 Euler 回路，并称图 G 为 Euler 图(欧拉图)。

直观地讲，Euler 图就是从一顶点出发且每边恰通过一次能回到出发点的那种图，即不重复地行遍所有的边再回到出发点。在引入本章的哥尼斯堡七桥问题中，欧拉证明了不存在这样的回路，使它经过图中每条边且只经过一次又回到起始点。

定义 4.20 度数：图中连接某个顶点的边的数目称为该顶点的度数。度数为奇数的顶点叫做奇数度顶点；度数为偶数的顶点叫做偶数度顶点。

和欧拉图有关的结论参看下面的定理。

定理 4.4 若图 G 有奇数度顶点，则 G 的奇数度顶点必是偶数个。

定理 4.5 对连通图 $G(V,E)$， G 为欧拉图的充分必要条件是 G 的每一个顶点的度数都是偶数。

应用定理 4.5 去判断一个连通图是否为欧拉图比较容易，但要找出欧拉回路，当连通图比较复杂时就不太容易了，这里给出一种求欧拉回路的弗罗莱(Fleury)算法：①任取起始点记为 $v_0, v_0 \to R$；②设路 $R=\{e_1(v_0,v_{i_1}),\{e_2(v_{i_1},v_{i_2}),\cdots,\{e_r(v_{i_{r-1}},v_{i_r})\}$ 已选出，则从 $E\setminus\{e_1,e_2,\cdots,e_r\}$ 中选出边 e_{r+1}，使 e_{r+1} 与 v_{i_r} 相连，除非没有其他选择， $G\setminus\{e_{r+1}\}$ 仍应为连通的；③重复步骤②，直至不能进行下去为止。

对于中国邮递员问题来说，如果给出的街道图 G 是欧拉图，则很容易由弗罗莱算法求出一个欧拉回路。但是若 G 不是欧拉图，即存在奇度数的节点，则中国邮递员问题的解决要困难得多，根据定理 4.4，G 的奇数度节点必是偶数个，因而可将奇数度顶点分为若干对，每对顶点之间在 G 中有相应的最短路，将这些最短路画在一起构成一个附加的边子集 E'，令 $G'=G\cup E'$，即把附加边子集 E' 叠加在原图 G 上形成一个多重图 G'，这时 G' 中连接两个顶点之间的边不止一条，显然 G' 是一个欧拉图，因而可以求出 G' 的欧拉回路。

G' 的欧拉回路不仅通过原图 G 中每条边，而且还通过 E' 中的每条边，且均仅一次。中国邮递员问题的难点在于当 G 的奇数度节点较多时，可能有很多种配对方法，关键在于怎样选择配对，能使相应的附加边子集 E' 的权数 $w(E')$ 为最小？为此有以下定理。

定理 4.6 设 $G(V,E)$ 为一个连通的赋权图，则使附加边子集 E' 的权数 $w(E')$ 为最小的充分必要条件是 $G\cup E'$ 中任意边至多重复一次，且 $G\cup E'$ 中的任意回路中重复边的权数之和不大于该回路总权数的一半。

2. 奇偶点图上作业法

基于定理 4.6，可归结出搜寻中国邮递员问题最优解的算法——奇偶点图上作业法，现举例如下。

【例 4.18】已知邮递员要投递的街道如图 4.26(a)所示，试求最优邮路。

解：先找出奇节点：A_1，A_2，A_3，A_4，B_1，B_2，B_3，B_4，在奇节点进行配对，不妨把 A_1 与 B_1，A_2 与 B_2，A_3 与 B_3，A_4 与 B_4 配对。求其最短路，添加重复边，如图 4.26(b)所示，显然它不是最优解。

1) 第一次调整

删去多于一条的重复边，即 A_3 与 B_3，A_4 与 B_4 中的(A_4, B_3)，调整后，实际上成为 A_1 与 B_1，A_2 与 B_2，A_3 与 A_4，B_3 与 B_4 的配对，它们的最短路如图 4.26(c)所示。

2) 第二次调整

发现在回路{A_1, A_2, B_2, A_4, B_3, B_4, B_1, A_1}中重复边的权数和为 11，大于该回路权数 20 的一半，因而调整时，把该回路的重复边删去，代之以重复其余部分，得图 4.26(d)。可以

北京大学出版社版权所有 禁止转载

看出，实际上是调整为 A_1 与 A_2，B_1 与 B_4，A_3 与 A_4，B_2 与 B_3 配对。

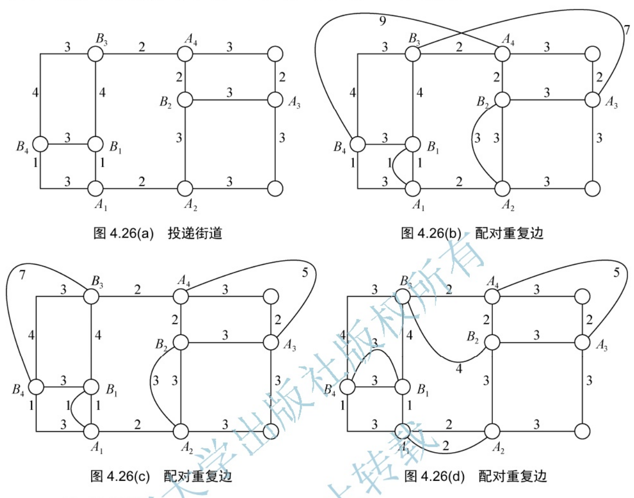


图 4.26(a) 投递街道

图 4.26(b) 配对重复边

图 4.26(c) 配对重复边

图 4.26(d) 配对重复边

3) 第三次调整

在图 4.26(d)中发现回路$\{A_3, A_4, B_3, B_2, A_3\}$中重复边的权数和为 9，大于该回路权数 14 的一半，因而删去原重复边(A_3, B_2)和(A_4, B_2)，而添加(B_2, A_3)和(A_4, B_3)，得到图 4.26(e)。

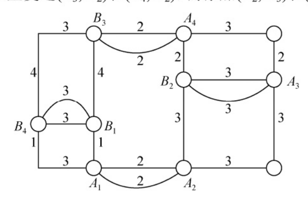


图 4.26(e) 配对重复边

进行检查发现，既没有多于一条的重复边，也没有任何回路使其重复边的权数之和大于该回路的一半，因此图 4.26(e)中的曲线部分就是最优的附加边子集 E'，而 $G \cup E'$ 为欧拉图，再由弗罗莱算法找出最优邮路。

在现实生活中，很多问题都可以转化为中国邮递员问题，例如道路清扫时如何使开空车的总时间最少的问题等。

北京大学出版社版权所有 禁止转载

4.5.2 旅行商问题

图论中还有一个看上去与欧拉回路问题很相似的著名问题，即哈密尔顿图问题。欧拉图问题考虑边的可行遍性，哈密尔顿图问题考虑点的可行遍性。

定义 4.21 Hamilton 圈：包含 G 的每个顶点的路叫做 Hamilton(哈密尔顿)路；闭的 Hamilton 路叫做 Hamilton 圈；含 Hamilton 圈的图叫做 Hamilton 图。

直观地讲，Hamilton 图就是从一顶点出发、每顶点恰通过一次能回到出发点的那种图，即不重复地行遍所有的顶点再回到出发点。

现讨论一个与哈密尔顿图问题有关的重要问题：一位旅行推销员想要访问 n 个城市中每个城市恰好一次，并且返回他的出发点。他应当以什么顺序访问这些城市以便旅行总距离最短？这就是著名的旅行商问题。这个问题可化归为以下的图论问题：设 G 是一个赋权完全图，各边的权非负，且有的边的权可能为 ∞ (对应两城市之间无交通路线)，求 G 中一条最短的哈密尔顿回路。

最直截了当的求解旅行商问题的方法是检查所有可能的哈密尔顿回路并且挑选出总权值最小的一条回路。若在图中有 n 个城市，则为了求解这个问题，得检查多少条回路？一旦选定了出发点，需要检查不同的哈密尔顿回路就有 $(n-1)!/2$ 条，因为第二个顶点有 $n-1$ 种选择，第三个顶点有 $n-2$ 种选择，以此类推。因为可以用相反顺序来经过一条哈密尔顿回路，所以最后的结果是 $(n-1)!/2$ 而不是 $(n-1)!$。

【例 4.19】考虑图 4.27(a)所示的赋权连通图，求出它的不同的哈密尔顿回路，并指出最短的哈密尔顿回路。

解：求哈密尔顿回路可以从任何顶点出发。下面给出从顶点 a 出发，考虑顺时针与逆时针顺序的不同的哈密尔顿回路。

$$G_1=(a,b,c,d,a) \qquad G_2=(a,b,d,c,a)$$
$$G_3=(a,c,b,d,a) \qquad G_4=(a,c,d,b,a)$$
$$G_5=(a,d,b,c,a) \qquad G_6=(a,d,c,b,a)$$

于是，当不考虑顺(逆)时针顺序时，可知 $G_1=G_6$，以 G_1 为代表，$w(G_1)=8$ [图 4.27(b)]；$G_2=G_4$，以 G_2 为代表，$w(G_2)=10$ [图 4.27(c)]；$G_3=G_5$，以 G_3 为代表，$w(G_3)=12$ [图 4.27(d)]。经过比较可知，G_1 是最短的哈密尔顿回路。

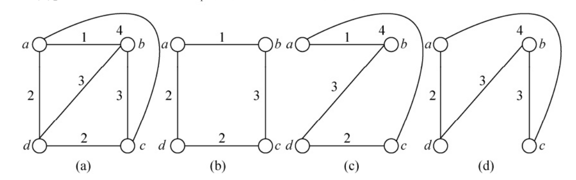


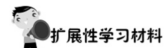
图 4.27 赋权连通图

不过，要注意的是，$(n-1)!/2$ 增长极快。当有几十个城市时，试图用这种方法来求解旅行商问题就不切实际。例如有 25 个城市，那么就不得不考虑总共 $24!/2$ (约 3.1×10^{23})条不同的哈密尔顿回路。假设检查每条哈密尔顿回路只花费 1ns(10^{-9} 秒)，那么就需要大约

北京大学出版社版权所有 禁止转载

1 000 万年才能求出这个图中长度最短的一条哈密尔顿回路。

“旅行商问题”是非常明确的，可是要找出一个行之有效且能迅速提供解答的方法，目前并不存在。在 1963 年美国的《管理科学》发表的一篇讨论“旅行商问题”的文章就说道：“人类由于他的运算能力的限制在解决旅行货郎问题上并不好。”人们现在对于这个问题的解决都是借助高速的电子计算机来运算，能够找到一些比较近似的结果。如改良圈算法就是先求一个 Hamilton 圈 C，然后适当修改 C 以得到具有较小权的另一个 Hamilton 圈，感兴趣的话可参考图论的有关文献。



车辆路径问题及其智能启发式算法

车辆路径问题(Vehicle Routing Problem，VRP)是运筹学中的热点问题。关于 VRP 的研究和应用最初来自于公共服务行业，如学生班车路径、银行运钞车路径等问题，后来随着电子商务和物流配送业的发展，VRP 也成为各商业企业、邮政快递、高速公路配货等行业亟待解决的问题。VRP 的原型是旅行商问题，即给定 n 个城市和两城市之间的距离，求一条访问各城市一次且仅一次的最短路线。VRP 问题是多路旅行商问题的扩展，每个城市(客户)有一个需求量，每个旅行商(车辆)有一定的载重量(能力)，一条路线上的客户总需求量不能超过行驶该路线的车辆的能力，求从仓库出发并回到仓库的一组总运输费用最少的路线。

解决 VRP 的方法大致历经 3 个时代：简单的启发式方法时代、启发式数学规划方法时代、精确优化算法与智能优化方法并存时代。较为成功的精确优化算法有 K 树法、拉格朗日放宽法等；而智能方法主要以模拟退火法、禁忌搜索法和遗传算法为代表。VRP 是一个 NP 完全问题，只有在客户数和路段较少时才有可能得到其精确的最优解，而一般情况下，只要找到一个可行的近优解就足够了，因此 VRP 的智能启发式算法成为人们研究的一个重要方向，这一时代的方法多数以 100 个客户的 Solomon VRPTW 标准问题集(6 类 56 问题)为算例进行验证，对比算法速度、规模、与已知最优解的偏差等，具有可比性。

北京大学出版社版权所有 禁止转载

本 章 小 结

图与网络的内容非常丰富，本章首先阐述了图的基本概念和图的矩阵表示，然后对最小支撑树、最短路径、最大流，以及最小费用最大流问题及其算法做了详细的讨论，最后就物流活动中的中国邮递员和旅行商问题做了简要的描述和分析。



关键术语(中英文)

图(Graph)	边(弧)(Edge(Arc))
连通图(Connected Graph)	链(路)(Chain(Path))
赋权图(Weighted Graph)	支撑子图(Spanning Graph)
破(避)圈法(Tear(Avoid) Cycle Method)	最短路(Shortest Path)
可行流(Feasible Flow)	增广路(Augmenting Path)
截集(Cut Set)	标号法(Labeling Method)

最大流(Maximum Flow)　　最小费用(Minimum Cost)
中国邮递员问题(Chinese Postman Problem)　　旅行商问题(Travel Salesman Problem)



知识链接

图论——欧拉

瑞士数学家，1707 年 4 月 15 日生于瑞士巴塞尔，1783 年 9 月 18 日卒于俄国圣彼得堡。他 15 岁在巴塞尔大学获学士学位，翌年获得硕士学位。1727 年，欧拉应圣彼得堡科学院的邀请到俄国。他以旺盛的精力投入研究，在俄国的 14 年中，他在分析学、数论和力学方面做了大量出色的工作。1741 年受普鲁士腓特烈大帝的邀请到柏林科学院工作，达 25 年之久。在柏林期间他的研究内容更加广泛，涉及行星运动、刚体运动、热力学、弹道学、人口学，这些工作和他的数学研究相互推动。欧拉这个时期在微分方程、曲面微分几何，以及其他数学领域的研究都是开创性的。



欧拉是 18 世纪数学界最杰出的人物之一，他不但在数学上作出伟大贡献，而且把数学用到了几乎整个物理领域。他又是一个多产作者。他写了大量的力学、分析学、几何学、变分法的课本，《无穷小分析引论》《微分学原理》《积分学原理》都成为数学中的经典著作。除了教科书外，他的全集有 74 卷。18 世纪中叶，欧拉和其他数学家在解决物理问题过程中，创立了微分方程这门学科。

欧拉引入了空间曲线的参数方程，给出了空间曲线曲率半径的解析表达式。1766 年他出版了《关于曲面上曲线的研究》，建立了曲面理论。这篇著作是欧拉对微分几何最重要的贡献，是微分几何发展史上的一个里程碑。欧拉在分析学上的贡献不胜枚举。如他引入了 Γ 函数和 B 函数，证明了椭圆积分的加法定理，最早引入了二重积分等。他还解决了著名的图论问题：柯尼斯堡七桥问题。在数学的许多分支中都常常见到以他的名字命名的重要常数、公式和定理。

北京大学出版社版权所有
禁止转载



习题 4

4.1 填空题

(1) 若用图来表示某排球单循环赛中各队的胜负情况，则可用点表示运动队，用__________表示它们间的比赛结果。

(2) 在一个网络中，流 f^* 是可行流，当且仅当满足__________和__________。

(3) 一个连通图中的最小树可能不唯一，其权是__________。

(4) 增广路上与发点到收点方向一致的边必须是非饱和边，方向相反的边必须是流量__________的边。

(5) 任一树中的边数和它的顶点数之间的关系是__________。

4.2 判断题

(1) 从起点出发到终点的最短路是唯一的。　( )

(2) 图论中的图，可以改变点与点的相互位置，只要不改变点与点的连接关系。( )

(3) 在任一无向图中，支撑树是含边数最少的连通图。　( )

(4) 若在网络图中不存在关于可行流 f 的增广路时，f 即为最大流。　( )

(5) 截集中弧的容量之和称为截量。　( )

4.3 已知美国有 5 个城市 Atlanta，Chicago，Cincinnati，Houston，Las Vegas，其距离矩阵见表 4-3，求连接 5 个城市最短距离的电话网络。

表 4-3 美国 5 个城市的距离矩阵

j \ i	Atlanta	Chicago	Cincinnati	Houston	Las Vegas
Atlanta	0	702	454	842	2 396
Chicago	702	0	324	1 093	2 196
Cincinnati	454	324	0	1 137	2 180
Houston	842	1 093	1 137	0	1 616
Lasvagas	2 396	2 196	2 180	1 616	0

4.4 图 4.28 是 6 个城市 $v_1, v_2, \cdots, v_6$ 之间的一个公路网，每条公路为图中的边，边上的权数表示该段公路的长度(单位：百公里)。设你处在城市 v_1，那么从 v_1 到 v_6 应选择哪一路径使你的费用最省。

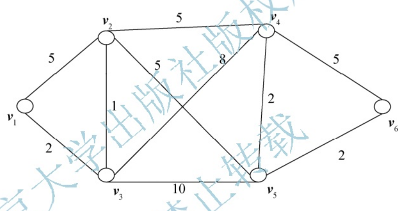


图 4.28 交通网络

北京大学出版社版权所有 禁止转载

4.5 某公司在 6 个城市 $c_1, \cdots, c_6$ 中有分公司，从 c_i 到 c_j 的直接航程票价记在下述矩阵中的 (i, j) 位置上(∞ 表示无直接航路)。帮助该公司设计一张任意两城市间的票价最便宜的路线。

$$
\begin{bmatrix}
0 & 50 & \infty & 40 & 25 & 10 \\
50 & 0 & 15 & 20 & \infty & 25 \\
\infty & 15 & 0 & 10 & 20 & \infty \\
40 & 20 & 10 & 0 & 0 & 25 \\
25 & \infty & 20 & 10 & 0 & 55 \\
10 & 25 & \infty & 25 & 55 & 0
\end{bmatrix}
$$

4.6 在图 4.29 中，(1)用 Dijkstra 方法求 v_1 到各点的最短路；(2)指出对 v_1 来说哪些顶点是不可到达的？

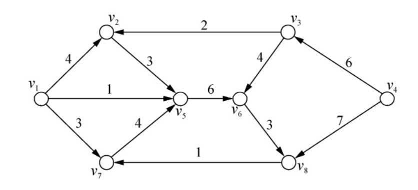


图 4.29　网络

4.7 在如图 4.30 所示的网络中，每弧旁的数字是 (c_{ij}, f_{ij})，试求：

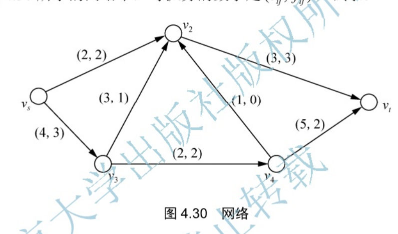


图 4.30　网络

北京大学出版社版权所有
禁止转载

(1) 确定所有的截集；

(2) 最小截集的容量；

(3) 说明给出的流是最大流。

4.8 某城市建设了一个从湖中抽水到城市的蓄水池的管道系统，如图 4.31 所示。线上标注的数字是单位时间通过两节点的容许流量，试求单位时间由湖到蓄水池的最大流量(单位：吨)。

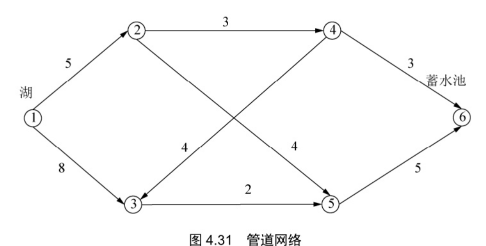


图 4.31　管道网络

4.9 求如图 4.32 所示网络的最小费用最大流，每弧旁的数字是 (b_{ij}, c_{ij})。

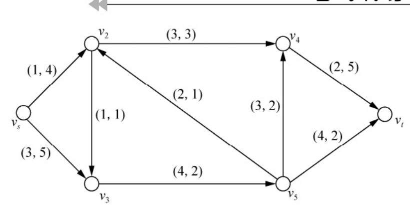


图 4.32　网络

4.10 求如图 4.33 所示的中国邮递员问题。

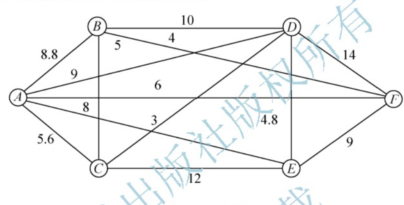


图 4.33　网络

北京大学出版社版权所有 禁止转载

4.11 各城市之间的航线距离见表 4-4。从北京(Pe)乘飞机到东京(T)、纽约(N)、墨西哥城(M)、伦敦(L)、巴黎(Pa)五城市旅游，每城市恰去一次再回北京，应如何安排旅游线，使旅程最短？

表 4-4　各城市之间的航线距离

j \ i	L	M	N	Pa	Pe	T
L		56	35	21	51	60
M	56		21	57	78	70
N	35	21		36	68	68
Pa	21	57	36		51	61
Pe	51	78	68	51		13
T	60	70	68	61	13	



已知某地区的配送网络如图 4.34 所示，其中点代表配送点，边代表公路，权为配送点间公路距离，问配送中心应建在哪个配送点，可使离配送中心最远的那个配送点所走的路程最近？如若还考虑配送点的运量，问配送中心应如何选择？

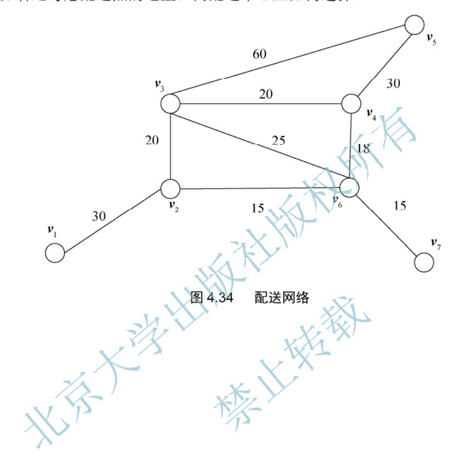


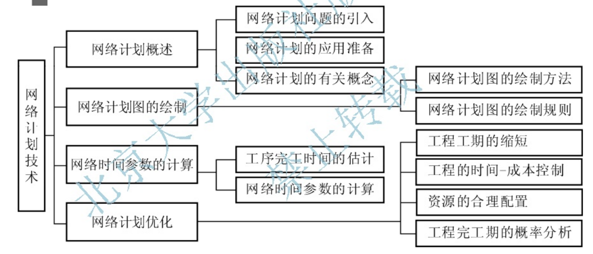
图 4.34　配送网络

北京大学出版社版权所有
禁止转载

第 5 章　网 络 计 划



【本章知识架构】



北京大学出版社版权所有
禁止转载

【本章教学目标与要求】

- 掌握网络计划图的有关概念，熟悉网络图绘制的规则，能依照作业明细表绘制网络图。
- 熟练掌握网络时间参数的计算方法，能利用网络时间参数确定关键路线。
- 熟练网络计划的时间优化方法；网络计划时间-成本优化方法；了解资源的合理配置方法。
- 理解 PERT 的工序工时的估计方法和工期的概率计算。



运储物流公司的配送作业流程

配送中心的物流成本的 60%左右发生在仓储系统，作业工序的不合理会造成大量人力和机器设备的耗费，作业工序衔接不畅也会造成配送时间的浪费，对物流作业时序的分析是配送目标准确完成的前提。

运储物流在进行作业时序优化时，首先把握时间和地域上服务的侧重点，了解客户需求多样化趋势，根据不同地域的不同客户对各种物资需求时间的不同，考虑物流作业时间安排；然后分析配送该货物所需的作业流程，各作业环节的平均作业时间，绘制物流作业工序流程图和物流作业时序表，接着找出关键路线；最后根据配送要求围绕配送中心进货、存储、拣货、补货、出货、配送等交叉和平行作业的整体运作要求。

通过作业分析，配送中心从功能布局、作业流程、货物储位三方面进行了整体优化，削减了 3 个作业环节，作业人数减少了 5 人，做到所有物流作业环节的关键路线的统筹优化。

资料来源：杨学春. 某物流企业仓库布局及作业流程分析与改进. 合肥学院学报，2014.

计划是管理四大职能之一，在管理活动中占有首要的地位。在现代社会复杂经济条件下，计划制订得好坏，往往有可能决定一项管理工作的成败，因而编制计划的方法也就显得越来越重要。网络计划技术是 20 世纪 50 年代末发展起来的一种编制大型工程进度计划的有效方法，网络计划技术的应用过程一般是绘制网络图、计算时间参数、优化和调整网络计划。

北京大学出版社版权所有 禁止转载

5.1 网络计划概述

用网络图编制的计划称为网络计划。网络计划有一系列优点，特别适用于生产技术复杂、工作项目多且联系紧密的一些跨部门的工作计划，便于对计划进行控制、管理、调整和优化，更清晰地了解工作之间相互联系与相互制约的逻辑关系，掌握关键工作和计划的全盘情况。

5.1.1 网络计划问题的引入

网络计划技术是编制大型工程进度计划的有效方法，1956 年美国杜邦公司和兰德公司运用网络计划技术来统筹各项工作，并找出编制与执行计划的关键路线，称为关键路线法，简称 CPM；1958 年美国海军特种计划局和洛克希德航空公司在规划和研制“北极星”导弹计划时，开发出了计划评审技术，简称 PERT；1962 年产生了能够体现随机因素作用的图形评审技术；1972 年美国 G. L. 穆勒等人，在图形评审技术的基础上提出了风险评审技术。这些技术方法均以网络描述工序及工序之间的关系。

网络计划技术最初是作为大规模开发研究的计划、管理方法而被开发出来的，但现在已被应用到世界军用、民用等各方面大大小小的项目中。在美国，已规定承包与军用有关的项目时，必须以 PERT 为基础提出预算和进度计划并取得批准，我国对网络计划技术的推广与应用也较早，1965 年著名数学家华罗庚教授首先在我国推广和应用了这些新的计划

管理方法，按照其主要特点——统筹安排，把这种网络计划技术称为“统筹法”。

【例 5.1】阿波罗登月计划(1958—1969)：阿波罗登月计划的全部任务分别由地面、空间和登月三部分组成，是一项复杂庞大的工程项目，它不仅涉及火箭技术、电力技术、冶金和化工等多种技术，为把人安全地送上月球，还需要了解宇宙空间的物理环境以及月球本身的构造和形状，它耗资 300 亿美元，研制零件有几百万种，共有两万家企业参与，涉及 42 万人，历时 11 年之久，为完成这项工作，除了考虑每个部门之间的配合和协调工作外，还要估计各种未知因素可能带来的种种影响，面对这些千头万绪的工作，千变万化的情况，就要求有一个总体规划部门运用一种科学的组织管理方法，综合考虑，统筹安排来解决。

【例 5.2】三军联合作战演习：空军要夺取制空权，对敌实施地面攻击，运送空降兵；海军舰艇护卫，运送陆军、海军陆战队登陆夺取滩头阵地，登陆完成后的巩固阵地与纵深发展；电子对抗部队实施情报收集分析与电子对抗。

参与兵种：海军航空兵、海军陆战队、水面舰艇部队、空军歼击机、攻击机、轰炸机、电子对抗机各团、坦克、炮兵、步兵、防化兵、通信兵、侦察兵、导弹部队等。需迅速制订科学的作战演习计划，以便对作战演习过程进行有效的管理与控制。

【例 5.3】大型工程实施(三峡工程、南水北调工程)有以下活动：产品设计、仿真、试制、中试；原材料设备订货、采购、运输、入库；厂房、设备施工建筑、安装；产品计划、生产、销售、安装、调试、维护。

参与单位涉及国家各部门、各行业、事业单位，为高速度、低成本、高质量，并在规定期限内完成该工程项目，其关键在抓好科学技术，抓好项目管理，组织协调好各单位、各任务、各工序的完成。

国内外应用网络计划的实践表明，网络计划技术是一种编制大型工程进度计划的有效方法，该技术既可以应用于新产品研制开发、大型工程项目、生产技术准备、设备大修等计划；又可以应用在人力、物力、财力等资源的安排，合理组织报表、文件流程等方面。

这里主要介绍两类网络计划技术——关键路线法和计划评审技术。关键路线法借助网络表示各项工作与其所需要的时间，以及各项工作的相互关系，通过网络分析研究工程费用与工期的相互关系，并找出在编制计划时及计划执行过程中的关键路线，而 PERT 方法注重于对各项工作安排的评价和审查。这两种方法在初期发展阶段的主要区别是：CPM 假设每项活动的作业时间是确定值，而 PERT 中作业时间是不确定的，是用概率方法进行估计的估算值；CPM 不仅考虑时间，而且还考虑费用，重点在于费用和成本的控制，而 PERT 主要用于含有大量不确定因素的大规模开发研究项目，重点在于时间控制；CPM 主要应用于以往在类似工程中已取得一定经验的承包工程，PERT 更多地应用于研究与开发。到后来两者有发展一致的趋势，以求得时间和费用的最佳控制。

5.1.2　网络计划技术的应用准备

网络计划一般用在复杂的项目管理中，网络计划必须做好下列准备工作。

1. 确定计划目标

网络计划的目标是多个方面和综合的目标，正是这个特点决定了网络计划的复杂性。一般按照不同的要求可以分成三类：一是时间要求；二是资源要求；三是费用要求。

2. 项目任务活动的分解

将整个项目根据技术上的需要分解为若干个互相独立的可执行的活动，通常把这种互相独立的活动称为工序或作业。

3. 确定各工序的衔接顺序

北京大学出版社版权所有 禁止转载

明确各项工序之间的先后逻辑关系。工序之间存在执行的先后顺序关系，如需先完成 *A* 工序才能进行 *B* 工序，则称 *A* 为 *B* 紧前工序，*B* 为 *A* 紧后工序，列出工序明细，以便建立整个项目的网络图以表示各项工序之间的相互关系。

4. 确定各个工序时间

确定各项工序所需的时间、人力、物力，一般将完成工序所需要的时间称为工时。

后面三项准备工作一般以工序明细表或作业明细表的形式给出，这些工作应该在管理活动中完成，本章默认这个工序明细表已经获得，所有分析都在这个前提下进行。

【例 5.4】某企业进行新产品研发的工序明细见表 5-1。

表 5-1　工序明细表

工　作	工作内容	紧前工序	工时/周	工　作	工作内容	紧前工序	工时/周
A	市场调研	/	4	*G*	制订生产计划	*F*	3
B	资金筹备	/	10	*H*	筹备设备	*B*，*G*	2
C	需求分析	*A*	3	*I*	筹备原材料	*B*，*G*	8
D	产品设计	*A*	6	*J*	安装设备	*H*	5
E	产品研制	*D*	8	*K*	调集人员	*G*	2
F	制订成本计划	*C*，*E*	2	*L*	准备开工投产	*I*，*J*，*K*	1

北京大学出版社版权所有 禁止转载

5.1.3　网络计划的有关概念

网络计划的重要标志是网络图，将项目中的所有活动之间的衔接关系表示出来。

1. 项目(Project)

也叫做工程，它是一项科学科研、建筑工程、生产任务，以及较复杂的工作。

2. 工序和事项

网络图中的箭线表示项目的工序；节点称为事项(事件)，代表一个工序开始或结束的时刻。某个工序的完成需要一定的时间与资源，事项不需要时间或所需时间忽略不计。

3. 网络图

是由工序和事项组成的具有一个发点和收点的有向赋权图，发点表示整个计划的开始，收点表示整个计划的结束。网络图有两种编制方法：一种是箭线法(双代号法)，用节点表示事项用箭线表示工序的网络图(Activity-on-Arrow，AOA)，它由工序、节点和路线三部分组成；一种是节点法(单代号法)，用箭线表示事项用节点表示工序的网络图(Activity-on-Node，AON)。根据需要，网络图可以分为总图、分图和工序流程图。

4. 路线

在网络图中，从发点(最初事项)到收点(最终事项)由各项工序衔接组成的一条有向路。在网络计划分析中，最关注的是消耗时间最多的路，也就是完成整个项目所需要的时间，这段时间叫做工期，并将这样的路叫做关键路线，关键路线上的工序称为关键工序。

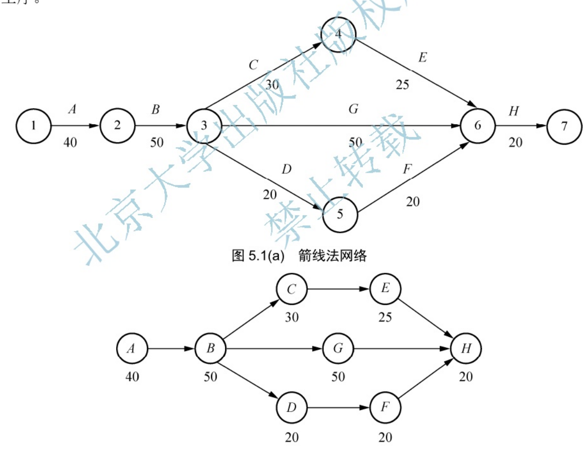
【例 5.5】某项目由 8 道工序组成，工序明细表见表 5-2。分别用箭线法和节点法绘制该项目的网络图，如图 5.1 所示。

表 5-2　工序明细表

序 号	代 号	工作名称	紧前工序	时间/天
1	*A*	基础工程	/	40
2	*B*	构建安装	*A*	50
3	*C*	屋面工程	*B*	30
4	*D*	专业工程	*B*	20
5	*E*	装修工程	*C*	25
6	*F*	地面工程	*D*	20
7	*G*	设备安装	*B*	50
8	*H*	试运行	*E*，*F*，*G*	20

解：箭线法网络图如图 5.1(a)所示，图中节点表示事项，如事项 4，表示工序 *C* 的完成，同时表示工序 *E* 的开始，描述了工序 *C* 和工序 *E* 的前后关系。只有当工序 *C* 完成后，工序 *E* 才能开始，工序 *C* 是工序 *E* 的紧前工序，当然工序 *E* 是工序 *C* 的紧后工序，同理，工序 *H* 的紧前工序是 *E*、*F*、*G*。

节点法网络图如图 5.1(b)所示，图中箭线描述了工序之间的紧前和紧后关系，节点表示工序。



北京大学出版社版权所有 禁止转载

图 5.1(a) 箭线法网络

图 5.1(b) 节点法网络

在图 5.1(a)中，从事项 1 到事项 7 有 3 条路线，不难看出，最长的路线为 1→2→3→4→6→7，其完成时间为 165 天，是整个项目的完工工期，这条路线也是该网络图的关键路线，工序 *A*、*B*、*C*、*E*、*H* 都是关键工序。

双代号 AOA 和单代号 AON 这两种网络图并没有本质的区别，各有其优缺点。AOA 在我国应用较多，比较难画，要引入虚工序；现实交流用 AOA 较多，因为 AOA 的表达方式更为清晰，手工较易分析，获得信息更多。相对而言，AON 在国外应用较多，更容易绘制和掌握，项目管理软件 Project 的网络图绘制原理就是 AON，特别在网络图发生一些改

变时较易调整。本章主要介绍双代号网络图的绘制和分析方法。

5.2 网络计划图的绘制

网络计划图是一种表示一项工程中各道工序的先后、衔接关系，以及所需要时间的图解模型。该种模型要求从整体的、系统的观点出发，全面地统筹项目中人、机、物等资源，并考虑各项活动之间相互依存的内在逻辑关系而绘制。

5.2.1 网络计划图的绘制方法

网络图的绘制一般都先画草图，再逐步调整。一般有两种方法：顺推法和逆推法。顺推法是指从项目距开工事项开始依次确定各道工序之间的先后顺序，并从开工事项画到完工事项为止；逆推法是指从工程完工事项开始依次确定各道工序之间的先后顺序，并从完工事项画到开工事项为止。

双代号网络图又称箭头图，由带箭头的线和点组成，箭线表示工作(或工序，活动)，节点表示事项。工作是完成整项任务的各项局部任务，需要一定的时间与资源，事项与工作相比，它不需要时间或所需要时间可以忽略不计。例如某工作 A 可以表示为图 5.2。

圆圈表示事项，里面的数字是各事项的编号，写在箭杆上中间的数字 15 为完成工作所需时间，表示工序 A 在事项点 3 开始，在事项点 5 结束，所消耗是时间是 15 个单位，该工序也表示为(3,5)，如图 5.2(a)所示。有时为了绘图的需要，还需引入虚工序，虚工序如图 5.2(b)所示，用来提示事项点之间的逻辑关系，虚工序不消耗资源，只是为了便于正确表示工序之间的先行后继关系。

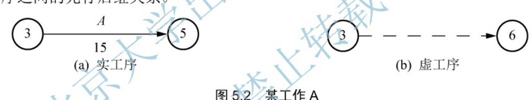


图 5.2 某工作 A

北京大学出版社版权所有 禁止转载

把表示各工序的箭线按照先后顺序以及逻辑关系，由左至右排列画成图；再给节点统一编号，按顺序地从小到大，由左向右；先编箭尾，后编箭头地编号，就得到网络计划图的雏形。网络计划图的绘制不能一蹴而就，要反复地修改和简化。

5.2.2 网络计划图的绘制规则

1. 绘制网络图需要遵守的规则

(1) 网络图只能有一个总开始事项，一个总结束事项。图 5.3(a)中网络图画法是错误的，因为有两个结束的事项 4、9。

(2) 网络图是有向图，不允许有回路，即不允许出现回路，否则就会发生逻辑错误，工作永远不能完成，图 5.3(b)是不允许的。

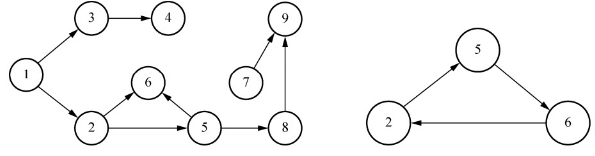


图 5.3(a) 错误的网络　　　　图 5.3(b) 错误的网络

(3) 事项 i, j 之间不允许有两个或两个以上的工序，图 5.3(c)是错误的，调整的办法引入一个新的事项，添加虚工序，如图 5.3(d)所示。

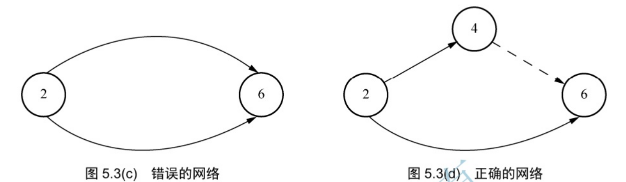


图 5.3(c) 错误的网络　　　　图 5.3(d) 正确的网络

(4) 必须正确表示工序之间的先行后继关系。如 4 道工序 A、B、C、D 的关系为：C 必须在 A、B 均完成之后才能开工，而 D 只需要在 B 完工后即可开工，画成图 5.3(e)是错误的，因本来与 A 工序无关的工序 D 被错误地标为必须在 A 完成后才能完工，添加虚工序修改为图 5.3(f)。

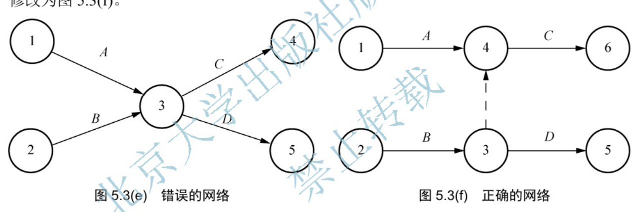


图 5.3(e) 错误的网络　　　　图 5.3(f) 正确的网络

北京大学出版社版权所有
禁止转载

(5) 虚工序的运用。虚工序在绘制网络图中运用很多，还可以用于正确表示平行工序与交叉工序。平行工序指在同样两个事项点之间进行的作业，交叉工序指一项作业不必全部完工才开始下一道工序，而是前道工序完成一部分，就可以继续下一道工序。平行和交叉工序在绘制网络图中要引入虚工序进行调整，在绘制网络图中要特别留意，如图 5.3(g)、(h)所示。

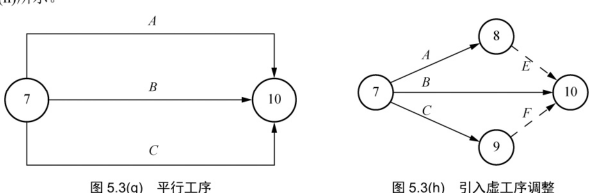


图 5.3(g) 平行工序　　　　图 5.3(h) 引入虚工序调整

两件或两件以上的工作交叉进行时，称为交叉工作。如作业 A 与作业 B 分别挖沟和埋

管子，那么它们的关系可以是挖一段埋一段，不必等沟全挖好后再埋，这就可以用交叉工作来表示，这时可以把这两件工作各分为几段，如 $A=A_1+A_2+A_3$，$B=B_1+B_2+B_3$，将这些作业拆分，就转变为上面熟悉的网络图表示方法，如图 5.3(i)所示。

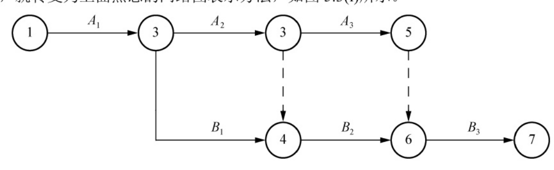


图 5.3(i)　作业拆分后的网络

2. 绘制网络图需要注意的问题

工序最好画成水平方式；最好能感觉出一条网络图的水平主线，可选择包含工序最多的一条路作为水平主线，然后调整其他工序的位置；给事项编号应在网络草图绘制完后进行，以便于阅读和计算。

【例 5.6】某项目由 10 道工序组成，工序明细表见表 5-3，试绘出网络图。

表 5-3　工序明细表

工序	工序代号	所需时间/(小时/天)	紧后工序	工序	工序代号	所需时间/(小时/天)	紧后工序
1	*A*	60	*B*，*C*，*D*，*E*	6	*F*	18	*J*
2	*B*	45	*J*	7	*G*	30	*I*
3	*C*	10	*F*	8	*H*	15	*J*
4	*D*	20	*G*，*H*	9	*I*	25	*J*
5	*E*	40	*H*	10	*J*	35	/

北京大学出版社版权所有 禁止转载

解：因为工序明细表给出的是紧后工序，故用逆推法绘制方便。为找出包含工序最多的路，对工序优先级进行编号。鉴于工序明细表给的是紧后工序，所以先对最后一道工序编号为 1，一直逆推，将其他工序的编号表示为——其所有紧后工序的编号中的最大值加 1。将找出的工序最多的路作为网络图的水平主线，先画草图，逐步调整，最后画出网络图，如图 5.4 所示。

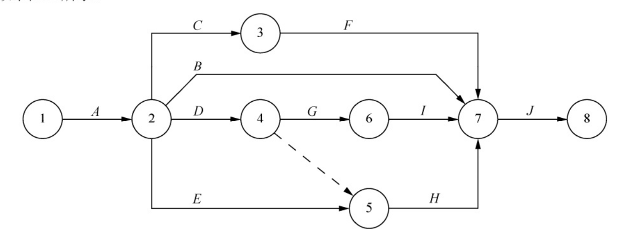




图 5.4　网络

5.3　网络时间参数的计算

网络图绘制完成后，下一步的工作就是计算网络时间参数，网络时间参数的计算的前提是估计出工序完工时间，它们用作业明细表的工时数据来估计。设网络图共有 n 个事项，记为 $1,2,\cdots,n$，其中 1 为项目开工事项，n 为项目完工事项，并用 $t(i,j)$或 t_{ij} 表示起点、终点事项分别是 i,j 工序完工时间。

5.3.1　工序完工时间的估计

前面提到，网络计划主要有 CPM 和 PERT 两种方法，它们的区别就在于工序完工时间的估计有所不同。CPM 采用单一时间计算，即各道工序的时间参数仅参考一个数值，估计时以完成工序的最大作业时间为准，通常就是作业明细表中的工时。

PERT 方法采用三时法计算工序完工时间，即预先估计 3 个时间，a 为最乐观时间，指顺利完成工序的最短时间；b 为最悲观时间，指考虑了最坏条件下完成工序的最长时间；m 为最可能时间，指正常条件下完成工序的最可能时间，三种时间发生的概率分别为1/6 、1/6 、4/6 ，则完工时间的期望值和方差为

$$t_{ij}=E(t_{ij})=\frac{a_{ij}+4m_{ij}+b_{ij}}{6}$$

$$\sigma_{ij}^2=D(t_{ij})=\left(\frac{b_{ij}-a_{ij}}{6}\right)^2$$

三时计算法较为简单，但估计较为粗糙，工序完成悲观和乐观时间是两个极端值，实际处理时可以对其值进行修正。例如最短时间 5～7 天的概率为 10%，最长时间 10～12 天的概率为 20%，正常时间为 9 天的概率为 70%，则期望值为$(6\times0.1+9\times0.7+11\times0.2)$天$=9.1$ 天。

【例 5.7】项目工序明细见表 5-4。

表 5-4　工序明细表

工序	紧前工序	工序的三种时间/天			工序	紧前工序	工序的三种时间/天		
		a	m	b			a	m	b
A	—	3	5	7	G	B，E	18	20	28

北京大学出版社版权所有 禁止转载

续表

工序	紧前工序	工序的三种时间/天			工序	紧前工序	工序的三种时间/天		
		a	*m*	*b*			*a*	*m*	*b*
B	—	8	9	16	*H*	*B*，*E*	26	33	52
C	—	8	11	14	*I*	*B*，*E*	18	25	32
D	*B*	2	4	6	*J*	*F*，*G*，*I*	12	15	18
E	*A*	3	4	5	*K*	*F*，*G*	11	21	25
F	*C*，*D*	8	16	18					

(1) 计算各工序完工时间的期望值和方差；

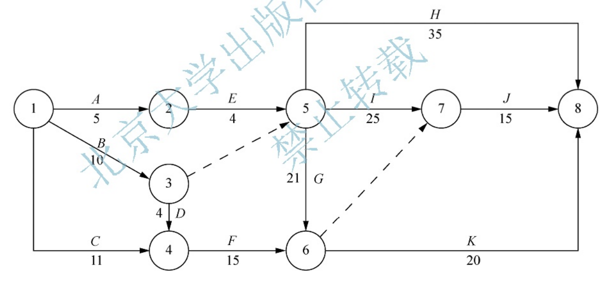
(2) 绘制项目的网络图。

解：(1) 根据三时计算公式，工序完工时间期望值和方差见表 5-5。

表 5-5　期望值和方差

工　序	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*
期望值/天	5	10	11	4	4	15	21	35	25	15	20
方　差	4/9	16/9	1	4/9	1/9	25/9	25/9	169/9	49/9	1	49/9

(2) 项目的网络图如图 5.6 所示，图中工序 *G*、*I* 是平行工序，必须添加一道虚工序。



北京大学出版社版权所有
禁止转载

图 5.5　项目的网络

5.3.2　网络时间参数的计算

网络时间参数主要包含事项时间参数和工序时间参数两大类，其中有最早开始时间、最迟结束时间和时差等时间参数。计算时都假设工程开始时间点为“0”，如 12 月 1 日开工，则将 12 月 1 日这一天记为第“0”天，而不是第 1 天。

1. 事项时间参数

事项时间参数有以下 3 个。

1) 最早开始时间 $t_E(j)$

$t_E(j)$ 指事项 j 的最早开始时间，是在保证该事项所有紧前工序都能够完成的前提下，从该事项开始的各项作业最早开始时间。

由于网络图是按照时间顺序安排的有向图，所以计算最早开始时间应选择工程开始的时间作为参照。计算方法是从初始事项开始，然后从左向右逐步计算各事项最早时间，直至终点事项，如图 5.6 所示，递推公式为

$$\begin{cases} t_E(1)=0 \\ t_E(j)=\max\limits_{i<j}\{t_E(i)+t(i,j)\} \qquad (j=1,2,\cdots,n) \end{cases}$$

$t_E(6)=5$ 6

$t_E(8)=\max\{t_E(6)+t(6, 8), t_E(7)+t(7, 8)\}$ 8

$t_E(7)=7$ 7

图 5.6 递推公式示意图

将计算出来的最早开始时间，写在各事项的指定位置，并用“□”括起来。

2) 最迟结束时间 $t_L(j)$

$t_L(j)$ 指事项 j 最迟结束时间，是在保证该事项所有紧后作业都不延误的前提下，该事项所有紧前作业的最迟结束时间。

最迟结束时间在计算时应选择项目结束的时间作为参照。计算方法是从结束事项开始，然后从右向左逐步计算各事项最迟结束时间，直至起点事项，如图 5.7 所示，递推公式为

$$\begin{cases} t_L(n)=t_E(n) \\ t_L(i)=\min\limits_{i<j}\{t_L(j)-t(i\ ,\ j)\} \qquad (i=1,2,\cdots,n-1) \end{cases}$$

将计算出来的最迟结束时间，写在各事项的指定位置，并用“△”括起来。

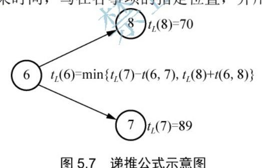


图 5.7 递推公式示意图

3) 事项的时差 $R(i)$

$$R(i)=t_L(i)-t_E(i) \quad i=1,2,\cdots,n$$

2. 工序时间参数

工序时间参数包括 6 个参数。

1) 工序 (i, j) 的最早开始时间 $t_{ES}(i, j)$

工序 (i, j) 的起点是事项 i，所以其最早开始时间应为事项 i 的最早时间，即

$$t_{ES}(i,j)=t_E(i)$$

在网络图上，工序的最早开始时间和事项的最早开始时间同样表示。

2) 工序 (i, j) 的最早结束时间 $t_{EF}(i, j)$

北京大学出版社版权所有 禁止转载

工序 (i, j) 的最早开始时间为 $t_{ES}(i, j)$，所以其最早结束时间为

$$t_{EF}(i, j) = t_{ES}(i, j) + t(i, j)$$

3) 工序 (i, j) 的最迟结束时间 $t_{LF}(i, j)$

工序 (i, j) 的终点是事项 j，所以其最迟结束时间应为事项 j 的最迟结束时间，即

$$t_{LF}(i, j) = t_L(j)$$

在网络图上，工序的最迟结束时间和事项的最迟结束时间同样表示。

4) 工序 (i, j) 的最迟开始时间 $t_{LS}(i, j)$

工序 (i, j) 的最迟结束时间为 $t_{LF}(i, j)$，所以其最迟开始时间为

$$t_{LS}(i, j) = t_{LF}(i, j) - t(i, j)$$

5) 工序总时差

是指在不影响总工期，即不影响其紧后作业最迟开始时间的前提下，工序最早开始时间可推迟开始的一段时间，记为 R_{ij} 或 $R(i, j)$，以表示工序安排上可以机动的时间。

$$R(i, j) = t_L(j) - t_{EF}(i, j) = t_{LF}(i, j) - t_{EF}(i, j) = t_{LS}(i, j) - t_{ES}(i, j)$$

即：总时差=最迟结束时间-最早结束时间=最迟开始时间-最早开始时间

6) 工序单时差

是指在不影响紧后作业最早开始时间前提下，从工序最早开始时间算起可推迟开始的时间，记为 r_{ij} 或 $r(i, j)$。

$$r(i, j) = t_E(j) - t_{EF}(i, j) = t_{ES}(j, k) - t_{EF}(i, j)$$

单时差等于其紧后工序的最早开始时间与本工序的最早结束时间之差，单时差也叫工序的自由时差，描述一项作业不影响各项紧后作业最早开工条件下单独拥有的机动时间，如图 5.8 所示。

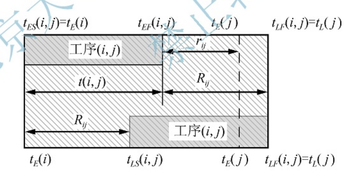


图 5.8　工序时间参数关系

将在网络图中总时差为零的工序称为关键工序，表明在时间安排上没有机动，由关键工序组成的从起点到终点的路线称为关键路线；最早开始时间与最迟结束时间相等的事项所连成的路线也是关键路线。

【例 5.8】按照期望作业时间计算例 5.7 网络图的时间参数，并指出关键路线。

解：首先计算网络图的事项时间参数。

1) 事项点的最早开始时间

从左向右逐步计算事项最早时间参数 $t_E(i)$，得

北京大学出版社版权所有 禁止转载

$$t_E(1)=0$$
$$t_E(2)=t_E(1)+t(1,2)=0+5=5$$
$$t_E(3)=t_E(1)+t(1,3)=0+10=10$$
$$t_E(4)=\max\begin{Bmatrix}t_E(1)+t(1,4)\\t_E(3)+t(3,4)\end{Bmatrix}=\max\begin{Bmatrix}0+11\\10+4\end{Bmatrix}=14$$
$$t_E(5)=\max\begin{Bmatrix}t_E(2)+t(2,5)\\t_E(3)+t(3,5)\end{Bmatrix}=\max\begin{Bmatrix}5+4\\10+0\end{Bmatrix}=10$$
$$t_E(6)=\max\begin{Bmatrix}t_E(5)+t(5,6)\\t_E(4)+t(4,6)\end{Bmatrix}=\max\begin{Bmatrix}10+21\\14+15\end{Bmatrix}=31$$
$$t_E(7)=\max\begin{Bmatrix}t_E(5)+t(5,7)\\t_E(6)+t(6,7)\end{Bmatrix}=\max\begin{Bmatrix}10+25\\31+0\end{Bmatrix}=35$$
$$t_E(8)=\max\begin{Bmatrix}t_E(5)+t(5,8)\\t_E(7)+t(7,8)\\t_E(6)+t(6,8)\end{Bmatrix}=\max\begin{Bmatrix}10+35\\35+15\\31+20\end{Bmatrix}=51$$

整个工程的完成时间，即工期为 51。

2) 事项的最迟结束时间

从右向左逐步计算事项最迟结束时间参数$t_L(i)$，得

$$t_L(8)=51$$
$$t_L(7)=t_L(8)-t(7,8)=51-15=36$$
$$t_L(6)=\min\begin{Bmatrix}t_L(8)-t(6,8)\\t_L(7)-t(6,7)\end{Bmatrix}=\min\begin{Bmatrix}51-20\\36-0\end{Bmatrix}=31$$
$$t_L(5)=\min\begin{Bmatrix}t_L(8)-t(5,8)\\t_L(7)-t(5,7)\\t_L(6)-t(5,6)\end{Bmatrix}=\min\begin{Bmatrix}51-35\\36-25\\31-21\end{Bmatrix}=10$$
$$t_L(4)=t_L(6)-t(4,6)=31-15=16$$
$$t_L(3)=\min\begin{Bmatrix}t_L(5)-t(3,5)\\t_L(4)-t(3,4)\end{Bmatrix}=\min\begin{Bmatrix}10-0\\16-4\end{Bmatrix}=10$$
$$t_L(2)=t_L(5)-t(2,5)=10-4=6$$
$$t_L(1)=\min\begin{Bmatrix}t_L(2)-t(1,2)\\t_L(3)-t(1,3)\\t_L(4)-t(1,4)\end{Bmatrix}=\min\begin{Bmatrix}7-5\\10-10\\16-11\end{Bmatrix}=0$$

将求出的事项时间参数填入网络图中，根据各事项的时差可以确定出关键路线，如图 5.9 所示，其中粗黑线标出的就是工程的关键路线，其上的事项都是时差为零的事项。

接着计算网络图的工序时间参数，一般在事项时间参数计算的基础上利用表格计算，见表 5-6，表中第二行表示各列的列标号及它们之间的关系，例如 4(2+3)表示第四列是第二列与第三列的和，8($3_{j\min}$－4)表示第八列是由第三列中元素 $3_{j\ \min}$ 和第四列对应元素的差，其中 $3_{j\min}$ 代表事项 j 的最早开始时间，可由第三列工序最早开始时间导出，关键工序是工序总时差为零的工序。

北京大学出版社版权所有 禁止转载

表 5-6　工序时间参数

工序 (i, j)	工时 $t(i, j)$	$t_{ES}(i, j)$	$t_{EF}(i, j)$	$t_{LF}(i, j)$	$t_{LS}(i, j)$	$R(i, j)$	$r(i, j)$	关键工序
1	2	3	4(2+3)	5	6(5−2)	7(5−4)	8($3_{j\min}$−4)	
$A(1, 2)$	5	0	5	6	1	1	0	
~~$B(1, 3)$~~	10	0	10	10	0	~~0~~	0	是
$C(1, 4)$	11	0	11	16	5	5	3	
$E(2, 5)$	4	5	9	10	6	1	1	
$D(3, 4)$	4	10	14	16	12	2	0	
~~(3, 5)~~	0	10	10	10	10	~~0~~	0	是
$F(4, 6)$	15	14	29	31	16	2	2	
~~$G(5, 6)$~~	21	10	31	31	10	~~0~~	0	是
$I(5, 7)$	25	10	35	36	11	1	0	
$H(5, 8)$	35	10	45	51	16	6	6	
(6, 7)	0	31	31	36	36	5	4	
~~$K(6, 8)$~~	20	31	51	51	31	~~0~~	0	是
$J(7, 8)$	15	35	50	51	36	1	1	

北京大学出版社版权所有
禁止转载

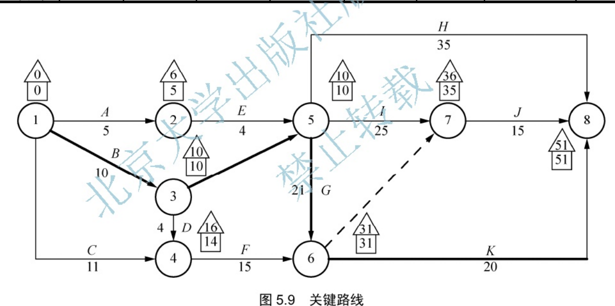


图 5.9　关键路线

最后得到关键工序和关键路线，总时差为零的工序为关键工序，$B(1, 3)$、$G(5, 6)$和 $K(6, 8)$都是关键工序；关键路线为 1→3→5→6→8，如图 5.9 中的粗黑线路。

5.4　网络计划的优化

网络技术不仅是绘制网络图并计算网络时间参数，得出一个初步的网络计划，其核心在于从工期、成本、资源等方面对这个初步计划做进一步的改善和调整，以求得最佳效果，该过程就是网络计划的优化。网络计划的优化，是在既定的条件下，初步拟定的网络计划方案，利用时差不断调整和改善，使之达到工期最短、成本最低、资源最优的目的。衡量网络计划是否达到最优，应综合评定工期、成本、资源消耗等技术经济指标，对某项工程

而言，如在技术资源有限的条件，希望施工进度最快，即资源有限、工期最短的优化；如既要保证按期完工，又要求投资最省，应寻求工期限定和成本最低的计划方案。

网络计划优化的内容有工期优化、费用优化和资源优化。工期优化的目的是使网络计划满足要求工期，保证按期完成工程任务，工期优化一般是通过调整关键工序的持续时间来满足工期要求的；费用优化寻求的目标是直接费用和间接费用最小时的工期，即最优工期；资源优化是在资源有限制的条件下如何使工期最短(资源有限、工期最短)，或在工期固定的条件下如何使资源均衡(工期最短、资源均衡)，资源优化的方法是通过改变工作的开始时间，使资源按时间的分布优化目标。

5.4.1 工程工期的缩短

缩短工程工期是网络计划要解决的重要问题之一，当然是在保证质量和现有资源允许的条件下，尽量缩短工程工期。具体包括以下主要途径。

(1) 压缩关键工序的作业时间。采取改进技术、工艺和设备等技术措施，优先保证关键工序所需要的人力与物力，组织力量对关键工序进行攻关，压缩关键工序的工序时间。

(2) 利用非关键工序的时差。非关键在开工时间上，具体工时上都有一定的弹性，因此从非关键工序上抽调部分人力、物力到关键工序上，以缩短关键工序的时间。

(3) 分解关键路线。重新将关键路线的作业分解，安排多工位或平行作业，以期压缩工作时间。

(4) 改变作业流程。在工艺流程允许的条件下，把关键路线上串联的关键工序改为平行工序或交叉工序，合理调配工程技术人员，缩短工期。(改变 PERT 网络图结构，工作量大，最后考虑。)

5.4.2 工程的时间-成本控制

在工程网络计划过程中，最关注的就是工期和成本，以使在工期和成本要求中达到一定的平衡。通过计算网络计划的不同完工期相应的总成本，以求得最低成本日程安排就是“工期-成本”控制要考虑的主要问题。

1. 有关成本和时间的概念

为便于讨论，将前面网络计划所述的工序时间称为工序的正常时间(Normal Time)，项目的完工期称为正常完工期，正常时间内完成工序的成本称为正常成本(Normal Cost)。当需要将完工期缩短到正常时间以下时，就要对原计划进行调整，缩短工序的时间，采取一些应急措施，如增加设备、雇用临时工、采用新技术等，这些措施必然伴随着增加成本，因采取应急措施而额外增加的成本加上正常成本称为工序的应急成本或赶工成本(Crash Cost)，工序时间不能无限制的缩短，工序的最短完成时间称为应急时间(Crash Time)。

工序时间的变化就会带来工程完工期的变化，缩短项目的完工时间虽然要增加应急成本，但同时也会增加收益，特别是时间的收益，比如提前完工获得的奖金、缩短时间而降低间接成本，有些项目提前完工就可以提前投产获得收益，有些公共项目还会获得更多的社会效益。就单个工序而言，缩短工序时间会增加成本，而对整个项目来说，缩短完工期有可能减少总成本。

工程的总成本=总正常成本+总应急增加成本-总应急效益=总应急成本-总应急收益

熟知了上面概念后，时间-成本控制研究的问题就比较具体了，即在采取应急措施条件下，考虑工程项目完工期为多少时总成本最低；给定项目缩短时间，如何调整计划使得总

北京大学出版社版权所有 禁止转载

成本最低；在不超过预算(总成本)的条件下，项目完工的最短时间是多少。

为分析这些问题，还需用到工序和工程的时间和成本的关系

$$单位时间工序的应急增加成本=\frac{应急成本-正常成本}{正常时间-应急时间}$$

这是指采取应急措施后，比正常施工时间增加的成本，也称成本斜率，用来衡量工序缩短一个时间单位所增加的成本。一般假设工序的成本斜率是一个常数，不发生变化，如图 5.10 所示。如某工序 (i, j) 正常工时为 5 天，成本为 600 元，按应急工时 3 天进行，所需成本为 900 元，则该工序的成本斜率 $c_{ij}=(900-600)元/(5-3)天=150$ 元/天，即每缩短一天需增加费用 150 元。

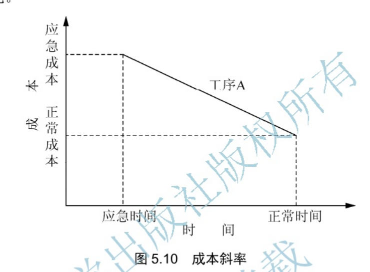


图 5.10 成本斜率

北京大学出版社版权所有
禁止转载

工程的边际成本用来衡量工程缩短或延长一个时间单位所带来的成本的变化，是考虑总成本变化的一个概念。

$$工程的边际成本=\frac{总成本的改变量}{提前或延长一个单位时间}$$

2. 时间-成本控制的应用

下面通过实例阐述时间-成本控制的分析方法。

【例 5.9】已知网络计划各项工作的正常时间，应急时间及相应费用见表 5-7。试：

(1) 绘制网络计划图，按照正常时间计算完成项目的总成本和工期；

(2) 按应急时间计算完成项目的总成本和工期；

(3) 按应急时间的项目完工期，调整计划使得总成本最低；

(4) 已知项目缩短一天额外获得奖金 5 万元，减少间接费用 1 万元，求总成本最低的项目完工期，也称为最低成本日程。

表 5-7 应急时间与费用

工序	紧前工序	时间/天		成本/万元		时间的最大压缩量	成本斜率/(万元/天)
		正常	应急	正常	应急		
A	/	19	15	52	80	4	7
B	*A*	21	19	62	90	2	14

续表

工序	紧前工序	时间/天		成本/万元		时间的最大压缩量	成本斜率/(万元/天)
		正常	应急	正常	应急		
C	B	24	22	24	30	2	3
D	B	25	23	38	60	2	11
E	B	26	24	18	26	2	4
F	C	25	23	88	102	2	7
G	D，E	28	23	19	39	5	4
H	F	23	23	30	30	0	+∞
I	G，H	27	26	40	55	1	15
J	I	18	14	17	21	4	1
K	I	35	30	25	35	5	2
L	J	28	25	30	60	3	10
M	K	30	26	45	57	4	3
N	L	25	20	18	28	5	2
总成本				506	713		

北京大学出版社版权所有 禁止转载

解：(1) 项目的网络计划图及其时间参数如图 5.11 所示，关键路线参看图中粗黑线。项目的完工期为 210 天，将表 5-7 正常成本一列相加，得到总成本为 506 万元。

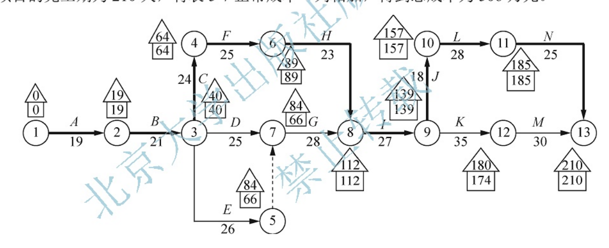


图 5.11 网络计划图及时间参数

(2) 项目网络图不变，时间参数如图 5.12 所示，项目的完工期为 187 天，将表 5-7 应急成本一列相加，得到总成本为 713 万元。

(3) 在第二个问题中，按应急时间项目的最早完成时间是 187 天，这段时间也叫应急工期，所有工序都按应急施工，总成本增加了 207 万元。实际上，非关键工序没有必要都按应急时间施工，图 5.12 中，非关键工序是 D、E、G、K 和 M，可以看到将工序 D、E、G 按正常时间施工时，都有机动时间，说明按正常时间施工也不一定影响项目完工期 187 天，如图 5.13(a)所示，工序 D、G 按正常时间施工完成总时间为 53 天，工序 E、G 按正常时间施工完成总时间为 54 天，而事项 3 和事项 8 之间的关键工序按应急时间完成的总时间为 68 天，所以非关键工序 D、E、G 有充裕的时间按正常时间完工，没有必要按应急时间赶工，从而增加不必要的应急成本；工序 K、M 按正常时间施工总共需要 65 天，按应急时间赶工需要 56 天，而事项 9 到事项 13 的关键工序按应急时间完工的总时间为 59 天，所以 K、M 也没有必要完全按照应急时间赶工，而只需缩短时间(65−59)天=6 天，不需缩短(65−56)天=9 天，如图 5.13(b)所示。

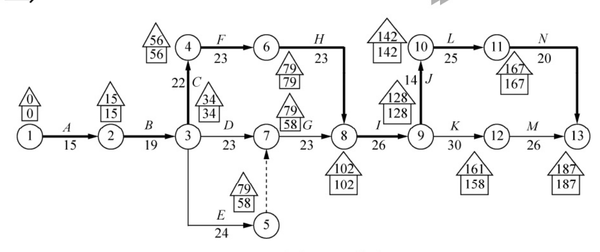


图 5.12　网络计划图及时间参数

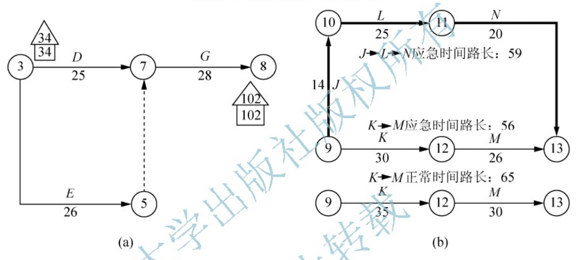


图 5.13　局部网络图

北京大学出版社版权所有 禁止转载

由表 5-7 可知，工序 K 缩短单位时间(1 天)的应急费用为 2，可以缩短 5 天；工序 M 缩短单位时间的应急费用为 3，可以缩短 4 天，这样缩短 6 天中安排是工序 K 缩短 5 天，工序 M 缩短 1 天，将使得增加的应急成本最小。按应急时间完成项目，最优的调整方案是：关键工序 A、B、C、F、H、I、J、L、N 全部按应急时间施工，工序 D、E、G 按正常时间施工，工序 K 缩短 5 天而 M 缩短 1 天，如图 5.14 所示。

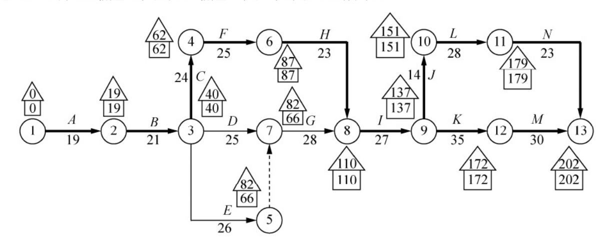


图 5.14　应急工期完成项目的最低总成本网络图

总成本等于正常成本加应急时间增加的成本之和，成本分析见表 5-8。

表 5-8 成本分析表

工 序	关键工序	正常时间	应急时间	实际使用时间	应急增加成本	正常成本	实际总成本
A	是	19	15	15	28	52	80
B	是	21	19	19	28	62	90
C	是	24	22	22	6	24	30
D		25	23	25	0	38	38
E		26	24	26	0	18	18
F	是	25	23	23	14	88	102
G		28	23	28	0	19	19
H	是	23	23	23	0	30	30
I	是	27	26	26	15	40	55
J	是	18	14	14	4	17	21
K	是	35	30	30	10	25	35
L	是	28	25	25	30	30	60
M	是	30	26	29	3	45	48
N	是	25	20	20	10	18	28
合 计					148	506	654

(4) 从问题(1)的计算可知，按正常时间施工的成本最低(506 万元)，如果按(3)成本增加 148 万元，缩短工期为 23 天。现在提前一天能获得收益 6 万元(奖金+间接费用)，不包括项目提前完成而获得的利润，工期缩短 23 天获得收益 138 万元，小于应急时间增加的成本，这里看能不能找到获得的收益不小于增加成本的方案。

首先考虑关键工序的时间。根据图 5.11 和表 5-7，选择应急一天增加成本小于或等于 6 的关键工序，从而缩短工程完工时间。这样的工序有 *C*、*J*、*N*，工序 *C* 可缩短两天，工序 *J* 缩短 4 天，工序 *N* 缩短 2 天。对图 5.11 进行调整得到图 5.15，得到两条关键路线，工序 *K* 和 *M* 变成关键工序，项目完工期为 202 天，缩短了 8 天，总成本变动额为

$$(2\times3+4\times1+2\times2-8\times6)\text{万元}=-34$$

总成本是否还能降低，只需检查图 5.15 中方框围起来的部分。要缩短工期必须两条关键路线同时缩短时间，上面一条线路工序 *N* 还能缩短 3 天，因此下面一条线路只对工序 *K* 缩短 3 天，对图 5.15 进行调整得到图 5.16，项目的完工期为 199 天，又缩短了 3 天，总成本变动额为

$$(3\times2+3\times2-3\times6)\text{万元}=-6$$

继续检查发现，缩短任何关键工序都不能减低成本，则总成本最低的项目工期是 199 天，总成本为(506−34−6)万元=466 万元，即最低成本日程为 199 天。

北京大学出版社版权所有 禁止转载

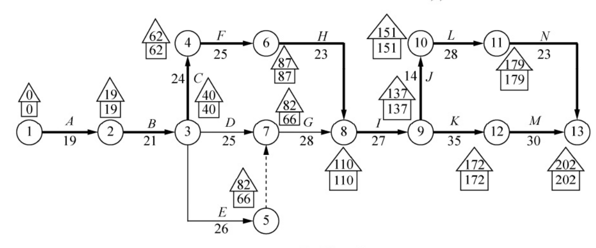


图 5.15　调整过程网络图

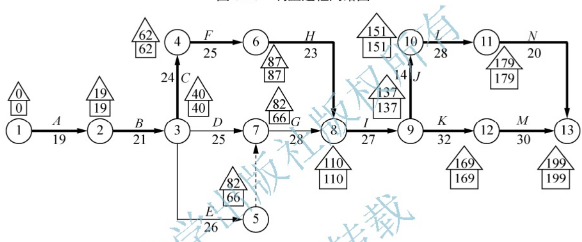


图 5.16　最低成本日程网络图

北京大学出版社版权所有
禁止转载

掌握了以上分析方法后，对于其他时间-成本控制问题都可以做类似的分析和求解。比如可将成本控制在某个范围内，对网络计划进行调整和优化，许多专用软件都设计了这种处理程序，为管理人员提供了更方便快捷的计划优化工具。

5.4.3　资源的合理配置

工程的可用资源在一般情况下都是有限的，完成工序的正常时间和正常成本都是在固定资源条件下估算出来的，如工序在正常情况下需要 15 人 5 台设备 20 天完成，总成本是 5 万元。在一定条件和范围内，工序的时间、资源和成本三者之间相互制约和相互转化。先进的施工设备能够节约工序时间，减少施工人员和工资，但同时也增加了设备投资。根据项目实施的实际情况和具体要求，因地制宜、因时制宜，对项目进行资源合理配置，也是网络优化的重要任务之一。

资源的合理配置大致包括几个方面：一是资源一定，如何组织、安排和调配资源保证项目按期完成；二是资源不足时，如何协调内部资源和采取应急措施保证项目按期完成；三是资源、时间和成本的整体调整和系统优化。这里仅通过一个人力资源安排的实例说明其原理。

【例 5.10】工程各工序的时间和资源需求见表 5-9。

表 5-9 工程各工序的时间和资源需求

工 序	紧前工序	每天需要人数	时间/天		成本/万元		压缩最大时间	成本斜率
			正 常	应 急	正 常	应 急		
A	—	5	10	8	30	70	2	20
B	*A*	12	8	6	130	150	2	10
C	*B*	20	10	7	100	130	3	10
D	*A*	12	7	6	40	50	1	10
E	*D*	20	10	8	50	80	2	15
F	*C*，*E*	10	3	3	60	60	0	0
G	*E*	7	13	9	70	86	4	4

(1) 绘制网络计划图，按正常时间计算项目完工期，按期完工最多需要多少人？

(2) 保证按期完工，如何采取应急措施，使得总成本最小又使得总人数最少，对网络计划进行优化分析。

解：(1) 绘制网络图，计算网络图的时间参数，如图 5.17 所示，项目完工期为 40 天，关键工序是 *A*、*D*、*E* 和 *G*，非关键工序是 *B*、*C* 和 *F*，它们的总时差都是 9，也是它们的全部机动时间。

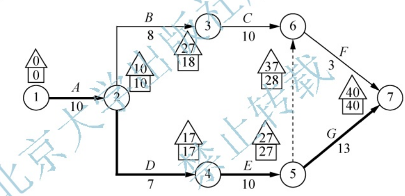


图 5.17 网络图

北京大学出版社版权所有 禁止转载

项目所需人数的计算方法首先安排关键工序的人员，然后对非关键路线的工序的开始时间进行调整，尽可能避开关键工序的用工高峰，最后各条路线上同一时间施工的最多人数就是项目所需人数。

为便于解释问题，对计划的资源配置使用时间坐标网络图更为方便，也叫资源负荷图。时间坐标网络图是网络图和横道图(甘特图)相结合的一种网络图，它兼具了网络图的逻辑性和横道图的直观性，横坐标为项目进度时间，纵坐标为工序，绘制方法如图 5.18 所示。

从图 5.18 可看出，如果非关键工序都按最早时间开始，第 1 天到第 10 天是用工低谷期，总共需 5 人；第 11 天到第 28 天是用工高峰期，第 19 天到第 27 天为 40 人，按此计划施工需要 40 人，显然这个方案很不经济。网络时间参数的计算为管理者提供了调整和优化网络计划的依据，减少施工人员就是利用非关键工序的时间差"削峰填谷"，调整非关键工序的开始时间，如果条件许可，某些非关键工序还可以分段作业、交叉作业，均衡利用资源。

将工序 *B* 按最早时间开始，工序 *C*、*F* 按最迟时间开始，调整后最多需要 32 人，如图 5.19 所示。

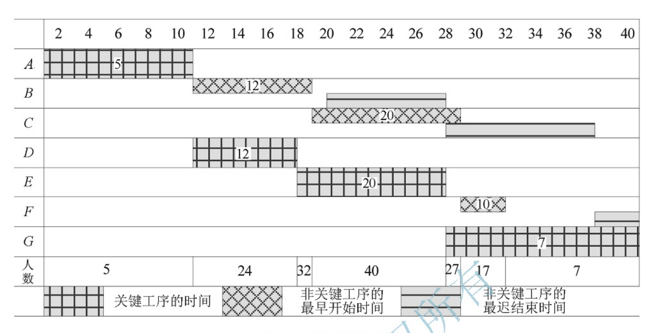


图 5.18　时间坐标网络图

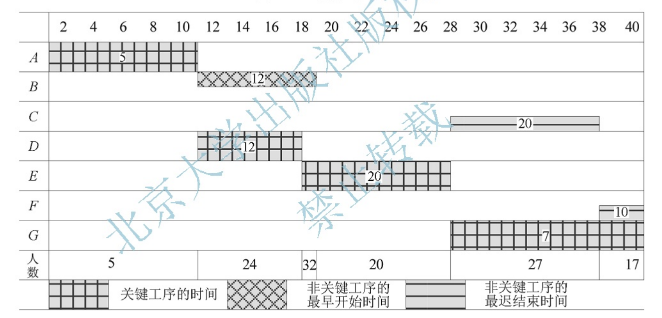


图 5.19　调整后的时间坐标网络图

北京大学出版社版权所有 禁止转载

(2) 从图 5.19 中可以看出，只有 1 天需要 32 人，对计划整体优化可考虑以下几个方案。

第一，对工序 B 或 E 采取应急措施，缩短工序时间 1 天，能够使总人数降到 27 人，从表 5-9 可知，工序 B 一天的应急成本比工序 E 低，所以工序 B 缩短 1 天，第 17 天完工，增加成本 10 万元。

第二，如果项目完工期推迟 1 天的成本比工序 B 的应急成本低，也可以考虑对关键工序 E 推迟 1 天开始，即第 19 天开始，项目完工期 41 天。

第三，从图 5.19 看出，人员并没有得到均衡利用，在某个时间段内就可以利用富裕的资源到关键工序，缩短关键工序时间，而在用工高峰期时将缩短的关键工序时间用到其他工序上。

第四，均衡利用资源，综合评价与审核。当资源、时间和成本相互可以替代时，制定转化标准，反复调整和优化网络计划，获得满意的计划方案。

网络计划技术不是单纯地编制网络图和计算时间参数，更重要的是对网络计划进行调整和优化，得到项目合理的决策方案。

5.4.4 工程完工期的概率分析

PERT 网络计划方法认为工序(i, j)的完成时间是一个随机变量，给出完成时间的 3 个估计值，最乐观估计a_{ij}、最悲观估计b_{ij}、最可能估计m_{ij}，则完成该工序的期望时间和方差(经验公式)为

$$Et(i,j)=\frac{a_{ij}+4m_{ij}+b_{ij}}{6},\quad \sigma^2_{t(i,j)}=\left(\frac{b_{ij}-a_{ij}}{6}\right)^2$$

一个工程的完工期为关键路线上各工序时间之和，所以完工期也是一个随机变量，由概率论极限定理可知，若关键工序比较多，工程的完工期就是一个服从正态分布的随机变量，其期望值T_E和方差σ_E^2都能从关键工序的期望和方差计算出来，即

$$T_E=\sum_{(i,j)\in I}\frac{a_{ij}+4m_{ij}+b_{ij}}{6}$$

$$\sigma_E^2=\sum_{(i,j)\in I}\left(\frac{b_{ij}-a_{ij}}{6}\right)^2$$

其中：I为关键路线上工序的集合，则完工期T服从正态分布$N(T_E,\sigma_E^2)$。

因而对于一个网络计划，只要算出关键路线上的期望完工时间和方差，就能对整个工程的完工时间进行概率计算。在计算时要将该正态分布转化为标准正态分布$N(0,1)$，然后查标准正态分布表就能得到在某个时间内完工的概率，当然也可在某个指定的概率下计算完工时间。

【例 5.11】已知一网络计划作业明细表及各个工序的a、m、b值(单位为月)，见表 5-10 的第 3、4、5 列。要求：

(1) 估计各工序的平均工时$t(i, j)$及标准差σ；

(2) 画出网络图，确定关键路线；

(3) 在 25 个月前完工的概率；

(4) 完工可能性达 95%的工期。

表 5-10 网络计划作业明细表

工　序	紧前工序	a	m	b	$t(i,j)$	σ	t_{ES}	t_{LS}	R	关键工序
A	/	7	8	9	8	0.33	0	3.33	3.33	
B	/	5	7	8	6.83	0.5	0	0	0	是
C	B	4	4	4	4	0	6.83	7.00	0.17	
D	C	3	4	6	4.17	0.5	10.83	11.00	0.17	
E	B	7	8	10	8.17	0.5	6.83	7.00	0.17	
F	A	6	9	12	9	1	8	11.33	3.33	
G	B	10	13	19	13.5	1.5	6.83	6.83	0	是

北京大学出版社版权所有 禁止转载

续表

工　序	紧前工序	a	m	b	$t(i,j)$	σ	t_{ES}	t_{LS}	R	关键工序
H	F，G，I	3	4	8	4.5	0.83	20.33	20.33	0	是
I	D，E	4	5	7	5.17	0.5	15	15.17	0.17	
J	D，E	7	9	11	9	0.67	15	15.83	0.83	

解：根据表 5-10 中的有关数据，估计各工序的平均工时。

(1) 根据三点时间法估计公式计算出各工序的平均工时 t 和方差 σ，填入表 5-10 第 6、7 列中。

(2) 根据表 5-10 所给的工序之间的相互衔接关系，绘制网络图，如图 5.20 所示。根据各工序的平均工时，计算各工序的网络时间参数，见表 5-10 第 8、9 和 10 列，从表中知总时差为 0 的工序为(1,3),(3,6),(6,7),所以关键路线为 1→3→6→7,总完成工期为 24.83(月)。

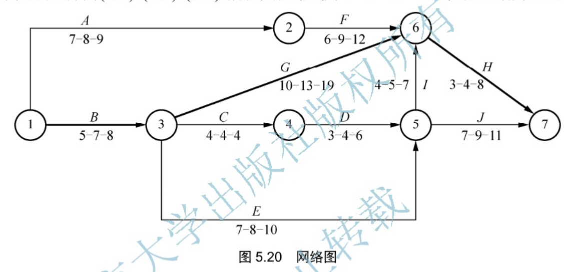


图 5.20　网络图

北京大学出版社版权所有
禁止转载

(3) 由于关键工序为 $B(1,3)$，$G(3,6)$，$H(6,7)$，所以

$$T_E = t(1,3) + t(3,6) + t(6,7) = 24.83$$

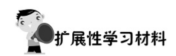
$$\sigma_E^2 = 0.25 + 2.25 + 0.69 = 3.19 = (1.79)^2$$

所以总工期服从正态分布 $N(24.83, 1.79^2)$，其在 25 个月前完工的概率为

$$P(T \leqslant 25) = \Phi\left(\frac{25 - 24.83}{1.79}\right) = \Phi(0.10)$$

查标准正态分布表得出 $\Phi(0.10)$=0.539 8，即此计划在 25 个月前完工的概率为 53.98%。

(4) 查标准正态分布表，知 $\Phi(1.64) = 0.95$，令 $\dfrac{T - 24.83}{1.79} = 1.64$，得 $T = 27.77$ 月，所以完工可能性达 95%的工期是 27.77 月。



图形评审技术 GERT

图形评审技术于 1962 年首先提出，又称决策网络技术或图示评审技术，是应用随机网络模型描述工程项目活动内容及其逻辑关系，以制订工程进度计划，并对其实现最优控制的一种计划管理方法。图解协

调技术的随机网络模型是由若干逻辑节点和连接两逻辑节点的有向支路构成的，可以对逻辑关系进行条件和概率处理(比如可能不执行某些活动)的一种网络分析技术。

图形评审技术与 CPM(Critical Path Method，关键路线法)相比，在网络逻辑和工作延续时间方面存在一定的概率陈述，即除了工序延续时间的不确定性外，还考虑工序存在概率分支，比如说，某些工作可能完全不被执行，某些工作可能仅执行其一部分，而另一些工作可能被重复执行多次。与 PERT(Program Evaluation and Review Technique，计划评审技术)相比，图形评审技术具有 4 个特点：①网络模型中的一些支路(即活动)不一定都能实现；②活动可以有多种结果；③网络图中有反馈环，即节点可以重复使用；④两节点间可以有两条以上支路，每一活动时间可有不同的概率分布。

GERT 多使用计算机仿真技术来模拟项目的执行情况，已成功地应用于空间技术研究、开发研究规划、存储分析、油井钻探、工业合同谈判、人口动态研究、系统可靠性研究、交通运输等领域。

本 章 小 结

网络计划始于把一个项目分成多个独立的活动，明确各个活动的前后顺序，估计每一个活动的完工时间，建立作业明细表。网络计划技术首先绘制网络图来表示所有这些信息，然后将这些信息转化为管理人员提供日程安排的信息，包括最早开始时间、最迟结束时间以及每一个活动的时差，确定出项目的关键路线和项目完工时间；最后为便于管理整个工程的进度和资源，给出了网络计划的优化方法，包括时间成本的平衡，资源的合理配置，以及在不确定状态下工程完工期的概率估计。

北京大学出版社版权所有 禁止转载



关键术语(中英文)

关键路线法(Critical Path Method)
三点估计法(Three Estimate Approach)
计划评审技术(Program Evaluation and Review Technique)
虚工序(Dummy Activity)
工序(Activity)
关键工序(Critical Activity)
节点(Node)
最早开始时间(Earliest Start Time)
事项(Event)
最迟开始时间(Latest Start Time)
交叉作业(Cross Job)
最早结束时间(Earliest Finish Time)
平行作业(Parallel Job)
最迟结束时间(Latest Finish Time)
紧前工序(Immediate Predecessor)
单时差(Single Slack Time)
紧后工序(Immediate Successor)
总时差(Total Slack Time)
图形评审技术(Graphic Evaluation and Review Technique)
风险评审技术(Venture Evaluation and Review Technique)


知识链接

统筹法——华罗庚

华罗庚教授出生于江苏省金坛县的一个贫寒之家，1924 年初中毕业后，在上海中华职业学校学习不到一年便因家贫失学，但他凭着强烈的求知欲和坚强的毅力，不管寒冬或酷暑，白天帮家里干活，晚上在油灯下刻苦攻读，孜孜不倦地钻研深奥的近代数学。功夫不负有心人，华罗庚于 1930 年在当时全国最重要的科技刊物《科学》上发表了关于代数方程式解法的文章，受到数学家熊庆来的重视，并被邀请到清华大学工作。起初，他做图书管理员，在此期间充分利用清华大学丰富的图书资料，如饥似渴地攻读一门又一门数学课程；后转做教学工作，并很快由助教升为讲师。1934 年，华罗庚成为中华教育文化基金会研究员，1936 年留学英国，在剑桥大学学习。在此期间，他连续发表了几篇有重要学术价值的学术论文，引起了世界数学界的注意。1938 年，华罗庚回到祖国，由于他的卓越成就，受聘为西南联合大学教授。此后他又应邀在苏联、美国的著名大学和研究机构任教。1950 年回国，先后任清华大学教授、中国科学院数学研究所所长和副院长等。



华罗庚教授一生成就辉煌，他在世界级刊物上发表过 150 多篇论文，写了 9 本书，其中有许多重要成果至今仍居世界领先水平。他还培养了一批中国数学界的骨干和年轻的新一代数学家，如段学复、闵嗣鹤、万哲先、王元、陈景润等。20 世纪 50—60 年代，根据中国国情和国际潮流，华罗庚教授积极倡导应用数学与计算机的研制；并亲自去全国各地普及应用数学知识与方法，为经济建设作出了巨大贡献。华罗庚教授的卓越成就，使他成为振兴中国近代数学的带头人和世界著名的第一流大数学家，他的名字与少数经典数学家一起被列入美国芝加哥科技博物馆等著名博物馆中。

北京大学出版社版权所有 禁止转载


习题 5

5.1 填空题

(1) 在网络图中，从始点出发，由各个关键活动连续相接，直到终点的线路称为__________。

(2) 在资源受限制时，时间与资源优化的方法之一，是先将有限的资源从__________作业调往__________作业，以便均衡地使用资源。

(3) 若用三种时间估计法计算作业时间，则应先估计出最乐观时间、最悲观时间和__________时间。

(4) 虚工序仅表示工序之间的逻辑关系，既__________消耗资源，也不占用时间。

(5) 某工程的各道工序已确定，为使其达到“成本最低、工期合理”的要求，进行优化时应采用__________技术。

5.2 判断题

(1) 从始点出发，经过连续相接的活动，直到终点的一条连线称为路线。 (　　)

(2) 网络图中任何一个节点都表示前一工序的结束和后一工序的开始。 (　　)

(3) 工序的总时差越大，则表明该工序在整个网络中的机动时间也越大。 (　　)

(4) 以同一节点为结束事项的各项作业最迟结束时间相同。 ( )

(5) 网络中通常只允许出现一条关键线路。 ( )

(6) 成本斜率越小，则每缩短单位作业时间所增加的费用就越小。 ( )

5.3 指出图 5.21 中所示的网络图的错误，并予以改正。

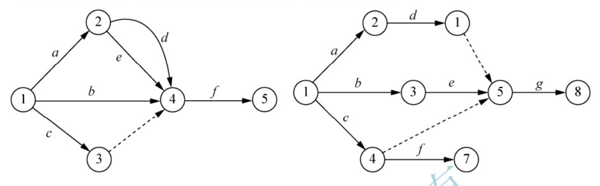


图 5.21 网络图

5.4 设某设备需进行一次大修，其各项活动的明细见表 5-11。

表 5-11 各项活动的明细

活动名称	紧前活动	工作时间/天
A	—	2
B	—	3
C	*A*	1
D	*B*	2
E	*A*	5
F	*C*，*E*，*D*	7
G	*C*	4

北京大学出版社版权所有 禁止转载

(1) 试编绘该设备大修理的网络图；

(2) 如果缩短活动 *E* 的工期，问是否会影响整个网络的工期？说明理由。

5.5 已知作业明细见表 5-12。

表 5-12 作业明细

工　序	紧前工序	工序时间/天	工　序	紧前工序	工序时间/天
A	—	10	*H*	*F*	10
B	—	8	*I*	*F*	4
C	*A*，*B*	6	*J*	*G*	12
D	*B*	16	*K*	*H*，*I*，*J*	16
E	*C*	24	*L*	*C*	8
F	*D*，*E*	4	*M*	*L*	24
G	*F*	4	*N*	*K*，*M*	4

要求：

(1) 绘制网络图；

(2) 计算各工序的最早开工时间、最早完工时间、最迟开工时间、最迟完工时间；

(3) 计算各工序的总时差；

(4) 确定关键路线。

5.6 某工程项目网络计划如图 5.22 所示，图中箭线下方为工序时间(月)。试确定该工程的完工时间，并按工序计算网络计划的时间参数，确定出关键路线。

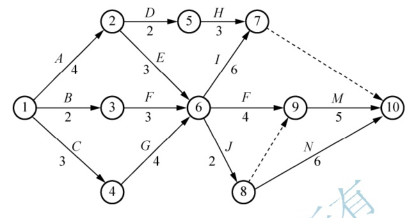
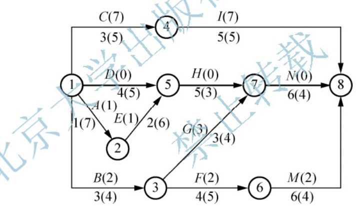


图 5.22 某工程项目网络计划

5.7 某项工程各道工序时间及每天需要的人力资源如图 5.23 所示。图中箭线上的英文字母表示工序代号，括号内数值是该工序总时差，箭线下左边数为工序工时，括号内为该工序每天需要的人力资源数。若人力资源限制每天只有 15 人，求此条件下工期最短的施工方案。



北京大学出版社版权所有
禁止转载

图 5.23 某项工程各道工序时间及每天需要的人力资源

5.8 某工程资料见表 5-13。

表 5-13 某工程资料

工　作	紧前工作	乐观时间 a	最可能时间 m	悲观时间 b
A	—	2	5	8
B	*A*	6	9	12
C	*A*	5	14	17
D	*B*	5	8	11
E	*C*，*D*	3	6	9
F	—	3	12	21
G	*E*，*F*	1	4	7

要求：(1)画出网络图；(2)求出每件工作工时的期望值和方差；(3)求出工程完工期的期望值和方差；(4)计算工程期望完工期提前 3 天的概率和推迟 5 天的概率。

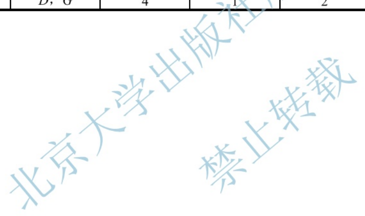


实际操作训练

某工程资料工序明细见表 5-14，时间单位为周，费用单位为万元。设每单位时间间接成本为 1 万元/周，求最低成本日程。

表 5-14 某工程资料工序明细

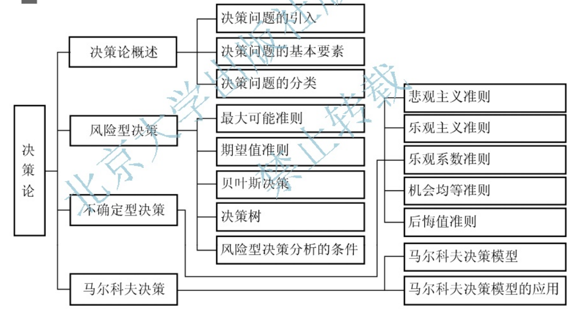
工　　作	紧前工作	正常时间	应急时间	正常费用	应急费用	成本斜率
A	—	6	3	4	5	1/3
B	—	5	1	3	5	0.5
C	*A*	7	5	4	10	3
D	*A*	5	2	3	6	1
E	*B*	6	2	4	7	0.75
F	*C*，*E*	6	4	3	6	1.5
G	*C*，*E*	9	5	6	11	1.25
H	*F*	2	1	2	4	2
I	*D*，*G*	4	1	2	5	1


北京大学出版社版权所有
禁止转载

第6章 决 策 论

北京大学出版社版权所有
禁止转载





【本章教学目标与要求】

- 掌握决策问题的基本要素，特别是决策问题的状态空间、策略空间、损益函数和决策准则；熟悉决策问题的分类。
- 掌握风险型决策的各种准则和常用的方法。
- 掌握不确定型决策的准则：悲观主义、乐观主义、乐观系数、机会均等和后悔值准则。
- 了解马尔科夫决策的特点及其适合解决的问题。



运储物流的智能配送决策

智能配送决策支持系统，是利用地理信息技术、多目标决策技术、路径优化模型、数据库技术等技术，依托高精度电子地图，对物流配送调度业务进行订单处理、优化分析、可视化调度报表输出、订单动态查询等，而建立的智能化、可视化的新型配送系统，旨在降低物流成本，提高客户服务水平，减轻调度人员和司机劳动强度，诸如满足电子商务、电话购物等现代物流配送业务的发展需要。

运储物流的配送决策系统以满足客户配送要求为前提，以车辆最少、里程最少、运输费用最低、时间最快、满意度最高等因素为目标，把配送订单科学地分配给可用的车辆，生成装车单和派车作业单，协同仓库部门一起配送任务。系统提供了配载订单的明细列表、装货顺序、车型、送货顺序、上下货时间窗、任务完成时间表等，为其物流配送业务提供有力的支持，使得配送业务透明可控，并降低总成本，提高客户满意度，为企业创造了持久的竞争力。

资料来源：http:// www.56888.net/news/20091221/266322415.html.

决策是管理永恒的主题，诺贝尔奖获得者赫伯特·西蒙认为：管理就是决策。决策是决策人(主观)对可行方案(客观)的价值识别过程，会受到各种决策环境、主观因素等的影响和干扰，幸运的是随着计算机通信技术的发展，在一定程度上替代了人们对一些常见问题的决策分析过程。

6.1 决策概述

北京大学出版社版权所有 禁止转载

决策是一项复杂的活动，依据所研究问题要达到的目标，按照一定的数量准则对可选方案进行比较和选择。决策问题是由决策者和决策域构成的，决策者是决策的主体，决策域由决策空间、状态空间和损益函数构成。

6.1.1 决策问题的引入

前面各章涉及决策问题的可选方案主要集中在某种确定环境上，例如线性规划问题在决策中都假定了从产品市场获得的利润是不变的，尽管也进行了一些灵敏度分析，但总体上考虑的仍然是一些微小的变化，目标函数的计算始终伴随着某些确定性的烙印，而现实当中的决策问题所面临的环境充满了变数。

在日常生活和企业组织的经营管理中，经常会遇到以下一些决策问题。

【例 6.1】宝洁、可口可乐公司开发的新产品进入市场。潜在顾客对新产品有什么反应？需投放多少新产品？在新产品投放整个市场之前是否要在某个区域进行试销？需投放多少广告费用配合新产品进入市场？

【例 6.2】IDG 风险投资公司保障安全投资。经济发展的政策环境如何？哪一个行业有最好的发展机会？现行利率是多少？这些因素对投资决策有哪些影响？

【例 6.3】政府部门招投标项目。投标项目的资质有哪些要求？标价如何设置？如何使招投标做到公开、公正？

【例 6.4】农业生产的合理安排。粮食在收割季节时的销售情况如何？未来季节的天气如何？投资成本为多少？

以上决策问题都面临着各种各样的不确定性，本章将建立具有这种不确定性的决策问题的分析框架，给出用数量方法合理选择决策方案的分析模式和方法。

6.1.2　决策问题的基本要素

看下面两个实例。

【例 6.5】某企业经过市场调查和预测得知，某新产品在今后 5 年中在市场上的销售为畅销、一般、滞销的概率分别 0.3、0.5 和 0.2。为使该新产品投产，该企业有两种可供选择的行动方案：一种方案是投资 16 万元新建一车间，按这种方案，市场畅销、一般和滞销三种情况下的利润情况分别为获利 50 万元、25 万元和亏损 5 万元；另一种方案是投资 3 万元扩建原有车间，在这种方案下，市场畅销、一般和滞销三种情况下的利润情况分别为获利 35 万元、20 万元和 5 万元，则该企业应确定哪一种行动方案较为合适？

【例 6.6】某一设备是影响武器装备系统效能的关键装备，必须在今后两年内配备部队 100 套。如果现在从友好国家进口设备，每套需要 450 万元。通过调查，依照目前国内的技术能力，在两年内研制成功的概率为 0.7，研制费用为 2 000 万元，若研制成功，生产费用为每套 200 万元，研制失败后再进口国外设备则费用为每套 500 万元。为了保证武器装备的配备，应当如何选择方案较为合适？

分析以上两个问题发现，这些决策问题都包括以下要素。

1. 自然状态

它描述了决策问题所处的各种状态，也叫状态空间。如投资问题有三种自然状态，即产品畅销、一般和滞销；武器装备问题有两种自然状态，即研制成功和研制失败，显然决策问题的自然状态是指一种客观存在。

2. 可供选择的行动方案

是指为解决决策问题，决策者可采取的行动。如投资问题，决策者可采取的行动方案有两种，即新建车间和扩建车间；武器装备问题，决策者可采取的行动有两种，即自行研制和进口。可供选择的行动方案的全体也叫策略空间。

3. 损益函数

是决策者采取了某一行动方案后可能产生的效果，也叫结局或后果。对于投资问题，若采取新建车间的方案，有三种后果，即产品畅销时获利 50 万元、销路一般时获利 25 万元和产品滞销时亏损 5 万元；若采取扩车间的方案，有三种后果，分别为畅销时获利 35 万元、销路一般时获利 20 万元和产品滞销时获利 5 万元。

由此可以看出，在上述决策问题中，由于自然状态的不确定性，不论决策者采取什么样的行动方案，都可能产生多种不同的结果，即决策后果具有不确定性。损益函数是在各种自然状态下对各种方案的一种价值评估，一般用损益表表出，其一般形式见表 6-1。

表 6-1　损益表

状　态 / 概　率 / 方　案	θ_1	θ_2	…	θ_n
	$P(\theta_1)$	$P(\theta_2)$	…	$P(\theta_n)$
d_1	u_{11}	u_{12}	…	u_{1n}
d_2	u_{21}	u_{22}	…	u_{2n}

北京大学出版社版权所有 禁止转载

续表

状态 / 概率 / 方案	θ_1	θ_2	…	θ_n
	$P(\theta_1)$	$P(\theta_2)$	…	$P(\theta_n)$
…	…	…	…	…
d_m	u_{m1}	u_{m2}	…	u_{mn}

其中 $d_i\ (i=1,2,\cdots,m)$ 表示可供选择的方案；$\theta_j\ (j=1,2,\cdots,n)$ 表示自然状态，根据经验知识，有时可能获得各种自然状态发生的概率 $P(\theta_j)$；$u_{ij}\ (i=1,2,\cdots,m;\ j=1,2,\cdots,n)$ 表示在自然状态 θ_j 选择方案 d_i 的收益值。

基于表 6-1 的形式，例 6.5 的损益表为表 6-2。

表 6-2 损益表

状态 / 概率 / 方案	畅销(θ_1)	一般(θ_2)	滞销(θ_3)
	$P(\theta_1)$ =0.3	$P(\theta_2)$ =0.5	$P(\theta_3)$ =0.2
投资新厂(d_1)	50	25	−5
扩建旧厂(d_2)	35	20	5

损益表是决策技术应用的信息基础，其信息从管理的实践过程中获得，是决策方法与管理实践的接口。

4. 决策准则

为了判定各个方案效果的好坏，应拟定相应的原则和标准，作为选择决策方案的准绳，这就是决策准则。对于不同类型的决策问题，应采用不同的准则。

对于确定型决策，由于其状态是确定的，故只要直接比较各方案的效果——损益值，即可评定方案的好坏。在决策分析中常使用条件值这个名词，所谓条件值，就是在一定条件(即某一自然状态)下，采取某一方案所产生的效果，可能为条件赢利、条件收益、条件费用、条件损失等。因此确定型决策问题的决策准则，也就是直接比较各策略的条件值。

对于不确定型决策，由于决策者不知道各自然状态发生的任何信息，因而其决策带有很强的主观性。选择什么样的准则，通常和决策人的心理因素有着密切联系。对于风险型决策，由于知道各自然状态发生的概率，故当采取某一方案时，则可算出相应于这一策略的期望收益(或损失)值，比较各策略的期望收益值，就可选定某一方案。

5. 效用

效用是指客观结果在决策者心中的价值，由于个人的心理特征不同，决策者对决策将会带来的结果的判断有着他个人独特的感觉和反应，这种感觉和反应就是决策者认为的效用。进行决策之前，决策者可能要进行客观结果与效用之间的转换，确定各种方案的各种可能结果的效用，选择效用最高的方案。这是由决策过程决定的，鉴于本书学习内容的特点，后面不再涉及。

6.1.3 决策问题的分类

决策所要解决的问题是多种多样的。按决策所依据的目标个数可分为：单目标决策与

北京大学出版社版权所有 禁止转载

多目标决策。按不同准则划分成各种决策问题类型。按决策者所面临的自然状态的确定与否可分为：确定型决策、风险型决策(又称统计型或随机型)和不确定型决策，其中风险型决策是本章讨论的重点。

1. 确定型决策

研究环境条件为确定情况下的决策。如某工厂每种产品的销售量已知，研究生产哪几种产品获利最大，它的结果是确定的。确定型决策问题通常存在着一个确定的自然状态和决策者希望达到的一个确定目标(收益较大或损失较小)，以及可供决策者选择的多个行动方案，并且不同的决策方案可计算出确定的收益值。这种问题可以用数学规划，包括线性规划、非线性规划、动态规划等方法求得最优解。但许多决策问题不一定追求最优解，只要能达到满意解即可。

2. 风险型决策

研究环境条件不确定，但以某种概率出现的决策。风险型决策问题通常存在着多个可以用概率事先估算出来的自然状态，以及决策者的一个确定目标和多个行动方案，并且可以计算出这些方案在不同状态下的收益值。

3. 不确定型决策

研究环境条件不确定，可能出现不同的情况(事件)，而情况出现的概率也无法估计的决策。这时，在特定情况下的收益是已知的，也可用损益表表示。

北京大学出版社版权所有
禁止转载

6.2 风险型决策

在风险型决策中，决策者在进行决策时并不确切知道哪一个事件(自然状态)将来一定发生，而只是根据已有的经验、资料和信息，设定或推算出各事件发生的概率，并据此进行决策。风险型决策准则包括最大可能准则、期望收益最大准则和期望机会损失最小准则，通常用的决策方法有损益表法和决策树法。

6.2.1 最大可能准则

从各状态中选择一个概率最大的状态来进行决策(因为一个事件，其概率越大，发生的可能性就越大)。在解决风险型决策问题时，选择一个概率最大的自然状态，把它看成是将要发生的唯一确定的状态，而把其他概率较小的自然状态忽略，这样就可以通过比较各行动方案在那个最大概率的自然状态下的损益值进行决策。实质上是将风险型决策问题当作确定型决策问题来对待，经常当某一自然状态发生的概率比其他状态发生的概率大得多，而相应的损益值相差不大时，可采用该准则来决策。

【例 6.7】用最大可能准则求解表 6-3 所描述的风险型决策问题。

表 6-3　损益表

农作物/(千元/公顷) \ 概率 \ 天气类型	极旱年	旱年	平年	湿润年	极湿年
	0.1	0.2	0.4	0.2	0.1
水稻	10	12.6	18	20	22
小麦	25	21	17	12	8
大豆	12	17	23	17	11
燕麦	11.8	13	17	19	21

解：由表 6-3 可知，“极旱年”“旱年”“平年”“湿润年”“极湿年”5 种自然状态发生的概率分别为 0.1、0.2、0.4、0.2、0.1，显然“平年”天气状态的概率最大。按照最大可能准则，可以将“平年”的发生看成是必然事件，而在“平年”状态下，各行动方案的收益分别是：水稻为 18 千元/公顷，小麦为 17 千元/公顷，大豆为 23 千元/公顷，燕麦为 17 千元/公顷。因种植大豆的收益最大，所以农场管理者选择种植大豆为最佳种植方案。

6.2.2　期望值准则

期望值准则[①]就是把每个方案的损益值视为离散型随机变量的取值，求出它们的期望值，并以此作为方案比较的依据，选择平均收益最大或者平均损失最小的方案作为最佳决策方案。期望值准则包含最大期望收益和最小(期望)机会损失两种决策准则。

1. 最大期望收益准则(Expected Monetary Value，EMV)

应用最大期望收益准则要求对应的损益表中各元素代表每对“方案-自然状态”的收益值。

首先计算各方案损益值的期望值：

$$\mathrm{EMV}(d_i)=\mathrm{E}(d_i)=\sum_{j=1}^{n}P(\theta_j)u_{ij}$$

式中：$\mathrm{E}(d_i)$——第 i 个方案损益值的期望值；u_{ij}——第 i 个方案在第 j 种状态下的损益值；$P(\theta_j)$——第 j 种状态发生的概率。

然后从这些期望收益值中选取最大者，并以它所对应的方案为决策方案，即

$$\max_{1\leqslant i\leqslant m}\{\mathrm{E}(d_i)\}\rightarrow d_k$$

① 期望本质上是一个加权平均数，在实际问题中具有重要的参考价值，可用均值的回归解释其重要作用。众所周知每个学生在一次考试中的表现既取决于他的能力也取决于他在考试当天的心情所引起的随机误差，教师是不知道每个学生的具体能力的，但他知道平均水平的学生是可以在 100 分中得到 70 分的。如果一个学生只得了 40 分，那么教师对他的能力的评估应是多少呢？这一评估不应当是 40 分。低于平均分 30 分之多的成绩可能是 a、b 两个因素形成的结果：a. 该学生的能力低于平均水平；b. 该学生在考试当天心情不佳。只有当心情是完全不重要之时，教师才能把 40 分作为他的评估。更为可能的是，能力和运气都会有所影响，因此教师的最佳猜测应是：该学生的能力固然低于平均水平，但也的确是运气不佳。最佳评估应为 40～70 分，这既反映了能力的影响，也反映了运气的影响。有统计实验表明，那些只得了 40 分的学生中，有一半多的人在下一次考试中有望超过 40 分。由于这些表现不佳的学生们的分数是趋于向着均值 70 分上升的，故而这一现象被称为“对均值的回归”(Regression to the Mean)，均值的回归是一个有着贝叶斯解释的古老统计学思想。

北京大学出版社版权所有　禁止转载

【例 6.8】某商业企业销售一种新产品，每箱成本 80 元，销售单价 100 元，如果商品当天卖不出去，就会因变质而失去其使用价值。目前对这种新产品的市场需求情况不是十分了解，但有去年同期类似产品的日销量资料可供参考，见表 6-4。现在要确定一个使企业获利最大的日进货量的方案。

表 6-4 日销量表

日 销 量	完成天数	概 率
25 箱	20	0.1
26 箱	60	0.3
27 箱	100	0.5
28 箱	20	0.1
总计	200	1.0

解：根据去年同类产品的销售情况，可知产品的市场自然状态(日需求量)为 25 箱、26 箱、27 箱、28 箱，可行进货方案也应在这四种情况中选择，计算出各种自然状态下各方案的损益值，填入表 6-5。

北京大学出版社版权所有 禁止转载

表 6-5 损益表

状态 / 概率 / 方案	25 箱	26 箱	27 箱	28 箱	期望值
	0.1	0.3	0.5	0.1	
25 箱	500	500	500	500	500
26 箱	420	520	520	520	510
27 箱	340	440	540	540	490
28 箱	260	360	460	560	420

根据题知条件，损益值的计算如下。

(1) 若方案为日进货量 25 箱，则在 4 种自然状态下，假如该日能卖出 25 箱，则每箱毛利(100−80)=20 元，25 箱共 25 元×20=500 元；假如该日能卖出 26 箱，但是只进 25 箱。所以，收益只有 500 元，在 27 箱和 28 箱的自然状态下同理。

(2) 若方案为日进货 27 箱，则在 4 种自然状态下，卖出 25 箱，收益 500 元，但是损失两箱，80 元×2=160 元，所以，收益只有(500−160)元=340 元；卖出 26 箱，收益 520 元，但是损失一箱，(520−80)元=440 元；卖出 27 箱，收益 540 元，假如该日能卖出 28 箱，但是只进 27 箱，收益只有 540。

(3) 如上法逐一计算，就可以将 4 种方案在各种自然状态下的损益值算出来，填入表 6-5 中。

为选择最优方案，计算各方案的期望值(EMV)：

$$EMV(25)=(0.1\times500+0.3\times500+0.5\times500+0.1\times500)\text{元}=500\text{ 元}$$
$$EMV(26)=(0.1\times420+0.3\times520+0.5\times520+0.1\times520)\text{元}=510\text{ 元}$$
$$EMV(27)=(0.1\times340+0.3\times440+0.5\times540+0.1\times540)\text{元}=490\text{ 元}$$
$$EMV(28)=(0.1\times260+0.3\times360+0.5\times460+0.1\times560)\text{元}=420\text{ 元}$$

根据最大期望收益准则，选择期望值最大的 510 元所对应的计划方案，即每天进货 26

箱为决策方案。

2. 最小(期望)机会损失准则(Expected Opportunity Loss，EOL)

应用最小(期望)机会损失准则要求对应的损益表中各元素代表每对“方案-自然状态”的机会损失值。

首先计算各方案的期望机会损失值为

$$\mathrm{EOL}(d_i)=\sum_{j=1}^{n}P(\theta_j)u'_{ij}$$

式中：$\mathrm{EOL}(d_i)$——第 i 个方案期望机会损失值；u'_{ij}——第 i 个策略方案在第 j 种状态下的损失值；$P(\theta_j)$——第 j 种状态发生的概率。

然后从这些期望机会损失值中选取最小者，并以它所对应的方案为决策方案，即

$$\min_{1\leqslant i\leqslant m}\{\mathrm{EOL}(d_i)\}\rightarrow d_k$$

【例 6.9】应用最小(期望)机会损失决策准则确定例 6.8 的最优决策方案。

这里的销售损失，包括进货量高于市场需求量而折价处理的损失，也包括进货量低于市场需求量造成缺货而失去销售机会的利润损失。计算步骤与最大期望值准则相同，详细过程见表 6-6。

北京大学出版社版权所有 禁止转载

表 6-6 损益表

方案＼概率＼状态	25 箱	26 箱	27 箱	28 箱	期 望 值
	0.1	0.3	0.5	0.1	
25 箱	0	20	40	60	32
26 箱	80	0	20	40	22
27 箱	160	80	0	20	42
28 箱	240	160	80	0	112

解：根据已知条件，各方案的损失值计算方法如下。

进货 25 箱，如遇上市场销售状态为 25 箱，损失为 0，如遇上市场销售状态为 26 箱，损失(100−80)元×1=20 元，遇上市场销售状态为 27 箱，损失(100−80)元×2=40 元，遇上市场销售状态为 28 箱，损失(100−80)元×3=60 元，填入表 6-6。

进货 27 箱，如遇上市场销售状态为 25 箱，损失 80 元×(27−25)=160 元，同理遇上市场销售状态为 26 箱损失为 80 元，遇上市场销售状态为 27 箱，损失为 0，遇上市场销售状态为 28 箱，损失为(100−80)元×1=20 元，填入表 6-6，以此类推。

计算各方案的期望损失值：

EOL (25)=(0.1×0+0.3×20+0.5×40+0.1×60)元=32 元

EOL (26)=(0.1×80+0.3×0+0.5×20+0.1×40)元=22 元

EOL (27)=(0.1×160+0.3×80+0.5×0+0.1×20)元=42 元

EOL (28)=(0.1×240+0.3×160+0.5×80+0.1×0)元=112 元

比较各个方案的期望损失值，进货 26 箱的期望损失最小，所以选择进货 26 这个方案为行动方案，该结果与最大期望收益准则的决策结果一致。

3. EMV 与 EOL 决策准则的关系

本质上，EMV 与 EOL 决策准则是相同的。若设 u_{ij} 为决策矩阵的收益值，当市场所需的销售量(自然状态)等于所选策略的产量时，收益值最大，即在收益矩阵对角线上的值，都是其所在列中的最大者，于是机会损失值的求法见表 6-7。

表 6-7　机会损失值

方案 \ 状态	θ_1	θ_2	…	θ_n
方案 \ 概率	$P(\theta_1)$	$P(\theta_2)$	…	$P(\theta_n)$
d_1	$u_{11}-u_{11}$	$u_{22}-u_{12}$	…	$u_{nn}-u_{1n}$
d_2	$u_{11}-u_{21}$	$u_{22}-u_{22}$	…	$u_{nn}-u_{2n}$
…	…	…	…	…
d_n	$u_{11}-u_{n1}$	$u_{22}-u_{n2}$	…	$u_{nn}-u_{nn}$

第 i 个方案的机会损失为

$$\mathrm{EOL}(d_i)=\sum_{j=1}^{n}P(\theta_j)(u_{jj}-u_{ij})=\sum_{j=1}^{n}P(\theta_j)u_{jj}-\sum_{j=1}^{n}P(\theta_j)u_{ij}=\sum_{j=1}^{n}P(\theta_j)u_{jj}-\mathrm{EMV}(d_i)$$

故当 EMV 为最大时，EOL 便为最小，用这两个决策准则得到的决策结果是一致的。

上述决策都从损益表出发，实施了计算各可行方案的期望值、比较各方案的期望值、选择最优可行方案的决策过程，以后将这样的决策方法称为损益表法。

4. 风险型决策的灵敏度分析

对于风险型决策问题，其各个方案的期望损益值是对状态概率准确预测的基础上求得的。由于状态概率的预测会受到许多不可控因素的影响，因而基于状态概率预测结果的期望损益值也不可能同实际完全一致，会产生一定的误差。这样，就必须对可能产生的数据变动是否会影响最佳决策方案的选择进行分析，这就是灵敏度分析。

【例 6.10】某企业拟增加销售收入，现有两种方案可供选择：一是降低销售价格；二是增加广告费用。该企业管理者经过研究，运用期望值决策法编制出决策分析表，见表 6-8。由于市场情况极其复杂，它受许多不可控因素的影响，因而销售状态的概率可能会发生变化。试针对这种情况，进行灵敏度分析。

表 6-8　损益表

方案/万元 \ 状态	适　销	滞　销	期望值
方案/万元 \ 概率	0.7	0.3	
降低销售价格(d_1)/万元	500	−200	290
增加广告费用(d_2)/万元	300	−100	180

解：1) 以最大期望收益准则确定最佳方案

北京大学出版社版权所有 禁止转载

$$\mathrm{E}(d_1)=[0.7\times500+0.3\times(-200)]\text{万元}=290\text{万元}$$
$$\mathrm{E}(d_2)=[0.7\times300+0.3\times(-100)]\text{万元}=180\text{万元}$$

所以，根据最大期望准则，选择降低销售价格(d_1)为最佳行动方案。

2) 灵敏度分析

考虑市场销售状态中适销的概率由 0.7 变为 0.3 时，则两个方案的期望收益值的变化为

$$\mathrm{E}(d_1)=[0.3\times500+0.7\times(-200)]\text{万元}=10\text{ 万元}$$
$$\mathrm{E}(d_2)=[0.3\times300+0.7\times(-100)]\text{万元}=20\text{ 万元}$$

所以，在 0.7 与 0.3 之间一定存在一概率点 P，当适销状态的概率等于 P 时，降低销售价格与增加广告费用方案的期望收益值相等，P 称为转移概率。

令 $\quad P\times500+(1-P)\times(-200)=P\times300+(1-P)\times(-100)$

得 $\quad P=0.33$

所以，当 $P>0.33$ 时，降低销售价格(d_1)为最佳方案；当 $P<0.33$ 时，增加广告费用(d_2)为最佳方案；当 $P=0.33$ 时，两个方案没有差异。

6.2.3 贝叶斯决策

风险决策中，自然状态将来发生的概率的准确程度会直接影响到方案的期望收益值，进而会影响到最大期望收益值。为了提高决策的期望收益值，常要进一步补充新信息，通过获得的新信息修正原先对自然状态发生概率的估计值，并利用修正的概率重新进行决策。由于修正概率是利用概率论中的贝叶斯(Bayes)公式来计算，故称这种决策为贝叶斯决策。一般贝叶斯决策包括先验分析、预验分析和后验分析三步。

北京大学出版社版权所有 禁止转载

1. 先验分析

决策者首先根据历史统计资料及经验等估计各自然状态出现的概率，称为先验概率，设为 $P(\theta_j), j=1,2,\cdots,n$，然后估计出各个方案在不同状态下的损益值，见表 6-1。

根据期望值准则，计算各方案的期望收益为

$$\mathrm{E}(d_i)=\sum_{j=1}^{n}P(\theta_j)u_{ij}\qquad i=1,2,\cdots,m$$

然后根据最大期望收益准则选择出最优方案，同时得到相应的最优期望值，记为 EMV(先)，$\mathrm{EMV}^*(\text{先})=\max\limits_{1\leqslant i\leqslant m}\mathrm{E}(d_i)=\mathrm{E}(d_k)$。

2. 预验分析

为了更好地决策，在条件许可的情况下，往往需要进一步补充新的信息，这些新信息有助于修正原来对自然状态出现概率的估计值，比如可以通过调查、咨询得到，而为了获得补充信息需支付一定的费用，这就是这些补充信息的价值。

1) 完美信息及完美信息价值

在所有的信息中，完全可靠、准确无误的信息自然是最好的信息，即预报某自然状态在实际中必定出现的信息，称这种信息为完美信息。

在完美信息下决策所能获得的期望收益值称为具有完美信息的期望收益值，记为 EPPI。若完美信息预报自然状态 θ_j 出现，风险型决策问题就变为确定型决策问题，相应的收益值 u_{ij} 自然为评价方案 d_i 的数量指标，最优方案由 $\max\limits_{1\leqslant i\leqslant m}u_{ij}$ 确定，因此具有完美信息的期

望收益值为

$$\text{EPPI}=\sum_{j=1}^{n}P(\theta_j)\max_{1\leqslant i\leqslant m}u_{ij}$$

EPPI 与 EMV^*(先)之间的差额正是因为拥有完美信息而使决策期望收益值增加的部分，这里将其称为完美信息价值，记作 EVPI，即 $\text{EVPI}=\text{EPPI}-\text{EMV}^*$(先)，显然任何信息的价值都不会超过完美信息价值。

【例 6.11】某化工企业由于某项工艺不够好，产品成本较高。现计划改进该项工艺，取得新工艺有三种方案：一是自行研发；二是从国内引进；三是从国外引进。根据市场需求分析和估计，产品销路为好、一般和差三种可能，且发生的可能性分别为 0.3、0.5 和 0.2。该项工艺改进后会直接影响企业的赢利状况，其相应的损益值见表 6-9，计算所述问题完美信息的价值。

表 6-9　损益表　　　　单位：万元

状态 / 概率 / 方案	销路好	销路一般	销路差
	0.3	0.5	0.2
自行研发(d_1)	50	20	−15
国内引进(d_2)	25	23	12
国外引进(d_3)	12	12	12

解：计算三种方案的期望收益值，根据最大期望收益值准则得

$$\max_{1\leqslant i\leqslant 3}\text{E}(d_i)=\max\{22,21.4,12\}=22\text{万元}=\text{E}(d_1)$$

求解得知，先验分析中最大期望收益值为 22 万元，所以

$$\text{EMV}^*(\text{先})=\max_{1\leqslant i\leqslant 3}\text{E}(d_i)=\text{E}(d_1)=22\text{ 万元}$$

在完美信息下，若预报市场销路好，由 $\max\limits_{1\leqslant i\leqslant 3}u_{i1}=\max\{50,25,12\}=50$，选自行研制($d_1$)为最优方案；若预报市场销路一般，由 $\max\limits_{1\leqslant i\leqslant 3}u_{i2}=\max\{20,23,12\}=23$，选国内引进($d_2$)为最优方案；若预报市场销路差，由 $\max\limits_{1\leqslant i\leqslant 3}u_{i3}=\max\{-15,12,12\}=12$，选国内引进或国外引进($d_2$ 或 d_3)为最优方案。

具有完美信息的期望收益值为

$$\text{EPPI}=\sum_{j=1}^{3}p(\theta_j)\max_{1\leqslant i\leqslant 3}u_{ij}=(0.3\times 50+0.5\times 23+0.2\times 12)\text{万元}=28.9\text{ 万元}$$

因此完美信息价值为 $\text{EVPI}=\text{EPPI}-\text{EMV}^*$(先)=(28.9−22)万元=6.9 万元。

2) 预验分析

如上所述，若记获得补充新信息支付的费用为 C_z，若 C_z 超过了其所能提高的期望收益值，这样的补充信息是不合算的。预验分析就是在补充新信息前，首先对补充新信息是否合算进行分析，然后决定是否补充新信息。具体地说，就是首先估算出完美信息的价值 EVPI，并将其作为补充信息费用 C_z 的一个预算标准，如果补充信息费用 C_z 远低于这个标准，则可认为这种补充信息是合算的；如果补充信息费用 C_z 接近甚至超过了这个标准，则这种补充信息是不合算的。

3. 后验分析

北京大学出版社版权所有 禁止转载

后验分析就是根据补充新信息，对先验概率进行修正得到后验概率，进而在此基础上进行决策，同时计算出补充新信息的价值并最终作出正确的决策。后验分析的具体工作步骤包括补充新信息、计算修正(后验)概率、后验决策、计算补充新信息的价值和重新决策。

1) 补充新信息

补充新信息就是获知$e_1,e_2,\cdots,e_s$共s个状态中的哪一个状态将出现，同时通过相关资料获取实际出现自然状态θ_j而预报e_k的概率，即得条件概率为

$$P(e_k|\theta_j)，(k=1,2,\cdots,s;\ j=1,2,\cdots,n)$$

2) 修正概率

由先验概率$P(\theta_j)\ (j=1,2,\cdots,n)$及条件概率$P(e_k|\theta_j)\ (k=1,2,\cdots,s;\ j=1,2,\cdots,n)$，根据贝叶斯公式计算出修正概率，即后验概率为

$$P(\theta_j|e_k)=\frac{P(e_k|\theta_j)P(\theta_j)}{\sum_{j=1}^{n}P(e_k|\theta_j)P(\theta_j)}，(k=1,2,\cdots,s;\ j=1,2,\cdots,n)$$

上面的后验概率也可以在表格中计算，见表6-10。

北京大学出版社版权所有
禁止转载

表6-10 后验概率计算表

		自然状态和先验概率				
		θ_1	…	θ_j	…	θ_n
		$P(\theta_1)$	…	$P(\theta_j)$	…	$P(\theta_n)$
条件概率 $P(e_k\|\theta_j)$	e_1	$P(e_1\|\theta_1)$	…	$P(e_1\|\theta_j)$	…	$P(e_1\|\theta_n)$
	⋮	⋮	…	⋮	…	⋮
	e_i	$P(e_i\|\theta_1)$	…	$P(e_i\|\theta_j)$	…	$P(e_i\|\theta_n)$
	⋮	⋮	…	⋮	…	⋮
	e_s	$P(e_s\|\theta_1)$	…	$P(e_s\|\theta_j)$	…	$P(e_s\|\theta_n)$
乘积概率 $P(e_k\|\theta_j)P(\theta_j)$	e_1	$P(e_1\|\theta_1)P(\theta_1)$	…	$P(e_1\|\theta_j)P(\theta_j)$	…	$P(e_1\|\theta_n)P(\theta_n)$
	⋮	⋮	…	⋮	…	⋮
	e_i	$P(e_i\|\theta_1)P(\theta_1)$	…	$P(e_i\|\theta_j)P(\theta_j)$	…	$P(e_i\|\theta_n)P(\theta_n)$
	⋮	⋮	…	⋮	…	⋮
	e_s	$P(e_s\|\theta_1)P(\theta_1)$	…	$P(e_s\|\theta_j)P(\theta_j)$	…	$P(e_s\|\theta_n)P(\theta_n)$
后验概率 $P(\theta_j\|e_k)$	e_1	$P(\theta_1\|e_1)$	…	$P(\theta_j\|e_1)$	…	$P(\theta_n\|e_1)$
	⋮	⋮	…	⋮	…	⋮
	e_i	$P(\theta_1\|e_i)$	…	$P(\theta_j\|e_i)$	…	$P(\theta_n\|e_i)$
	⋮	⋮	…	⋮	…	⋮
	e_s	$P(\theta_1\|e_s)$	…	$P(\theta_j\|e_s)$	…	$P(\theta_n\|e_s)$

3) 后验决策

若补充新信息预报将出现状态e_k，则根据后验修正概率分布$P(\theta_j|e_k)\ (j=1,2,\cdots,n)$，计算各方案的期望收益，即

$$\mathrm{E}(d_i|e_k)=\sum_{j=1}^{n}P(\theta_j|e_k)u_{ij}\quad (i=1,2,\cdots,m)$$

依据最大期望收益准则进行决策，即

$$\max_{1\leqslant i\leqslant m} \mathrm{E}(d_i \mid e_k) = \mathrm{E}(d_{i_k} \mid e_k)$$

因此 d_{i_k} 为预报将出现状态 e_k 时的最优选择方案，相应的预报将出现状态 e_k 时的期望收益值为

$$\mathrm{E}(e_k) = \mathrm{E}(d_{i_k} \mid e_k) = \max_{1\leqslant i\leqslant m} \mathrm{E}(d_i \mid e_k)\text{，}\quad k=1,2,\cdots,s$$

4) 计算补充新信息的价值

根据补充新信息预报出现状态的概率分布 $P(e_k)$ ($k=1,2,\cdots,s$)计算出后验分析中的最大期望收益值，记为 EMV^*(后)，即

$$\mathrm{EMV}^*(\text{后}) = \sum_{k=1}^{s} P(e_k)\mathrm{E}(e_k)$$

显然获得新信息后增加的期望收益值为补充新信息的价值，即 EMV^*(后)− EMV^*(先)为补充信息的价值。

5) 重新决策

根据补充新信息的价值与获得新信息所付出的费用 C_z 比较决定是否补充新信息，即若 EMV^*(后)− EMV^*(先)> C_z 则选择补充新信息；若 EMV^*(后)− EMV^*(先)< C_z 则不补充新信息。

【例 6.12】某公司准备经营某种新产品，可以采取的行动有小批量生产、中批量生产和大批量生产。如果大批量生产，在畅销时可获利 100 万元，一般时可获利 30 万元，滞销时会亏损 60 万元；如果中批量生产，在畅销时可获利 50 万元，一般时可获利 40 万元，滞销时会亏损 20 万元；如果小批量生产，在畅销时可获利 10 万元，一般时可获利 9 万元，滞销时会亏损 6 万元。根据以往经验，同类产品为畅销、一般和滞销的概率分别是 0.2、0.5 和 0.3。现公司考虑委托专门机构对产品的销路进行预测，预测费用为 5 万元，预测结果为畅销、一般和滞销。过去资料显示其预测的准确程度见表 6-11。

北京大学出版社版权所有
禁止转载

表 6-11　先验概率

预测结果 \ 概率 \ 状态		畅销 θ_1	一般 θ_2	滞销 θ_3
		$P(\theta_1)$ =0.2	$P(\theta_2)$ =0.5	$P(\theta_3)$ =0.3
预测畅销 e_1	$P(e_1 \mid \theta_j)$	0.80	0.20	0.02
预测一般 e_2	$P(e_2 \mid \theta_j)$	0.15	0.70	0.08
预测滞销 e_3	$P(e_3 \mid \theta_j)$	0.05	0.10	0.90

问该公司应如何决策？

解：采用贝叶斯决策方法进行决策。

1) 先验分析

根据已知条件得出损益表，参见表 6-12。

表 6-12　损益表

状　　态	畅销 θ_1	一般 θ_2	滞销 θ_3



方案 \ 概率	$P(\theta_1)$ =0.2	$P(\theta_2)$ =0.5	$P(\theta_3)$ =0.3
大批量(d_1)/万元	100	30	−60
中批量(d_2)/万元	50	40	−20
小批量(d_3)/万元	10	9	−6

首先计算三个方案的期望收益值为

$$\mathrm{E}(d_1)=P(\theta_1)u_{11}+P(\theta_2)u_{12}+P(\theta_3)u_{13}=0.2\times100+0.5\times30+0.3\times(-60)=17$$

$$\mathrm{E}(d_2)=P(\theta_1)u_{21}+P(\theta_2)u_{22}+P(\theta_3)u_{23}=0.2\times50+0.5\times40+0.3\times(-20)=24$$

$$\mathrm{E}(d_3)=P(\theta_1)u_{31}+P(\theta_2)u_{32}+P(\theta_3)u_{33}=0.2\times10+0.5\times9+0.3\times(-6)=4.7$$

根据最大期望收益值准则，得

$$\max_{1\leqslant i\leqslant3}\mathrm{E}(d_i)=\max\{17,24,4.7\}=24=\mathrm{E}(d_2)$$

应选择方案d_2，即企业应中批量生产，相应的最大期望收益为EMV^*(先)=24(万元)。

2) 预验分析

具有完美信息的期望收益值为

$$\mathrm{EPPI}=\sum_{j=1}^{3}P(\theta_j)\max_{1\leqslant i\leqslant3}u_{ij}=[0.2\times100+0.5\times40+0.3\times(-6)]\text{万元}=38.2\text{ 万元}$$

完美信息的价值为

$$\mathrm{EVPI}=\mathrm{EPPI}-\mathrm{EMV}^*(\text{先})=(38.2-24)\text{万元}=14.2\text{ 万元}$$

因为信息资料费$C_z=5<14.2$，所以初步认为委托专门机构预测是合算的。

3) 后验分析

(1) 补充新信息。专门市场预测机构将提供三种市场预测状态：预测畅销(e_1)、预测一般(e_2)和预测滞销(e_3)，其预测的准确程度条件概率$P(e_k|\theta_j)$，见表6-11。

(2) 修正概率。由先验概率$P(\theta_j)(j=1,2,3)$及条件概率$P(e_k|\theta_j)$，应用贝叶斯公式计算出修正概率，即后验概率$P(\theta_j|e_k)$，见表6-13。

北京大学出版社版权所有 禁止转载

表6-13 后验概率计算表

		自然状态和先验概率			
		θ_1	θ_2	θ_3	
		$P(\theta_1)$ =0.2	$P(\theta_2)$ =0.5	$P(\theta_3)$ =0.3	
条件概率 $P(e_k\|\theta_j)$	e_1	0.80	0.20	0.02	
	e_2	0.15	0.70	0.08	
	e_3	0.05	0.10	0.90	
乘积概率 $P(e_k\|\theta_j)P(\theta_j)$	e_1	0.160	0.100	0.006	$P(e_1)=0.266$
	e_2	0.030	0.350	0.024	$P(e_2)=0.404$
	e_3	0.010	0.050	0.270	$P(e_3)=0.330$

续表

		自然状态和先验概率			
		θ_1	θ_2	θ_3	
		$P(\theta_1)$ =0.2	$P(\theta_2)$ =0.5	$P(\theta_3)$ =0.3	
后验概率 $P(\theta_j \vert e_k)$	e_1	0.160/0.266	0.100/0.266	0.006/0.266	
	e_2	0.030/0.404	0.350/0.404	0.024/0.404	
	e_3	0.010/0.330	0.050/0.330	0.270/0.330	

4) 后验决策

若专门机构预测市场将出现状态 e_1 (畅销)，则根据后验修正概率分布 $P(\theta_j|e_1)$ ($j=1,2,3$)，计算各方案的期望收益值，即

$$\mathrm{E}(d_1 \mid e_1)=\sum_{j=1}^{3}P(\theta_j \mid e_1)u_{1j}=[(0.16\times 100+0.10\times 30-0.006\times 60)/0.266]\text{万元}=70.08\text{ 万元}$$

$$\mathrm{E}(d_2 \mid e_1)=\sum_{j=1}^{3}P(\theta_j \mid e_1)u_{2j}=[(0.16\times 50+0.10\times 40-0.006\times 20)/0.266]\text{万元}=44.66\text{ 万元}$$

$$\mathrm{E}(d_3 \mid e_1)=\sum_{j=1}^{3}P(\theta_j \mid e_1)u_{3j}=[(0.16\times 10+0.10\times 9-0.006\times 6)/0.266]\text{万元}=9.26\text{ 万元}$$

依据最大期望收益准则进行决策，即

$$\max_{1\leqslant i\leqslant 3}\mathrm{E}(d_i \mid e_1)=\max\{70.08, 44.66, 9.26\}=70.08\text{万元}=\mathrm{E}(d_1 \mid e_1)$$

因此 d_1 (大批量生产)为预测市场将出现状态 e_1 (畅销)时的最优选择方案，相应的预测出现状态 e_1 (畅销)的期望收益值为

$$\mathrm{E}(e_1)=\mathrm{E}(d_1 \mid e_1)=70.08\text{ 万元}$$

同样，若专门机构预测市场将出现状态 e_2 (一般)，则各方案的期望收益值为

$$\mathrm{E}(d_1 \mid e_2)=\sum_{j=1}^{3}P(\theta_j \mid e_2)u_{1j}=[(0.03\times 100+0.35\times 30-0.024\times 60)/0.404]\text{万元}=29.85\text{ 万元}$$

$$\mathrm{E}(d_2 \mid e_2)=\sum_{j=1}^{3}P(\theta_j \mid e_2)u_{2j}=[(0.03\times 50+0.35\times 40-0.024\times 20)/0.404]\text{万元}=37.18\text{ 万元}$$

$$\mathrm{E}(d_3 \mid e_2)=\sum_{j=1}^{3}P(\theta_j \mid e_2)u_{3j}=[(0.03\times 10+0.35\times 9-0.024\times 6)/0.404]\text{万元}=8.18\text{ 万元}$$

依据最大期望收益准则进行决策，得

$$\max_{1\leqslant i\leqslant 3}\mathrm{E}(d_i \mid e_2)=\max\{29.85, 37.18, 8.18\}=37.18\text{万元}=\mathrm{E}(d_2 \mid e_2)$$

因此 d_2 (中批量生产)为预测市场将出现状态 e_2 (一般)时的最优选择方案，相应的预测出现状态 e_2 (一般)的期望收益值为

$$\mathrm{E}(e_2)=\mathrm{E}(d_2 \mid e_2)=37.18\text{ 万元}$$

若专门机构预测市场将出现状态 e_3 (滞销)，则各方案的期望收益值为

$$\mathrm{E}(d_1 \mid e_3)=\sum_{j=1}^{3}P(\theta_j \mid e_3)u_{1j}=[(0.01\times 100+0.05\times 30-0.27\times 60)/0.33]\text{万元}=-41.52\text{ 万元}$$

$$\mathrm{E}(d_2 \mid e_3)=\sum_{j=1}^{3}P(\theta_j \mid e_3)u_{2j}=[(0.01\times 50+0.05\times 40-0.27\times 20)/0.33]\text{万元}=-8.79\text{ 万元}$$

北京大学出版社版权所有
禁止转载

$$\mathrm{E}(d_3 \mid e_3)=\sum_{j=1}^{3}P(\theta_j \mid e_3)u_{3j}=[(0.01\times10+0.05\times9-0.27\times6)/0.33]万元=-3.24万元$$

依据最大期望收益值准则进行决策，即

$$\max_{1\leqslant i\leqslant 3}\mathrm{E}(d_i \mid e_3)=\max\{-41.52,-8.79,-3.24\}=-3.24万元=\mathrm{E}(d_3 \mid e_3)$$

因此 d_3(小批量生产)为预测市场将出现状态 e_3(滞销)时的最优选择方案，相应的预测出现状态 e_3(滞销)的期望收益值为

$$\mathrm{E}(e_3)=\mathrm{E}(d_3 \mid e_3)=-3.24万元$$

5) 计算补充新信息的价值

后验分析中的最大期望收益值为

$$\mathrm{EMV}^*(后)=\sum_{k=1}^{3}P(e_k)\mathrm{E}(e_k)=[0.266\times70.08+0.404\times37.18+0.330\times(-3.24)]万元=32.59万元$$

补充新信息的价值为 EMV^*(后)− EMV^*(先)=(32.59−24)万元=8.59 万元。

6) 重新决策

因为 EMV^*(后)− EMV^*(先)> C_z，因此应当选择专门预测机构做市场预测。

经贝叶斯决策分析，该公司委托专门机构做市场预测。若预测市场为畅销，则选择大批量生产；若预测市场为一般，则选择中批量生产；若预测市场为滞销，则选择小批量生产。

6.2.4 决策树

决策树是风险决策问题直观表示的图示法，因为图的形状像树，所以被称为决策树。当利用决策树法时，损益表表示的缺点均能被克服，同时决策树法还方便简捷、层次清楚，能形象地显示决策过程。

决策树是一种可用以帮助决策者进行决策的树状图，如图 6.1 所示，它由决策节点、方案分支、事件节点、概率分支和结果节点按一定关系连接而成。决策节点常用小方框(□)代表，表示需在此处进行决策，从它向后引出的每一分枝代表可能选取的一个策略或方案；事件节点用小圆圈(○)表示，从它引出的分枝代表其后继状态，分枝上标明的数字表明该状态发生的概率，常称这种分枝为概率枝；结果节点为决策树的末梢，用小三角形(△)表示，代表决策问题的一个可能结果，旁边的数字为这种情况下的损益值，如图 6.1 所示。

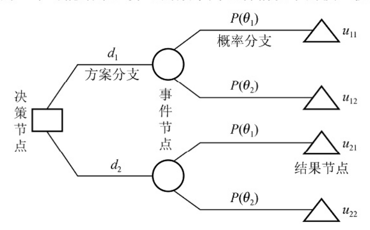


图 6.1 决策树符号示意图

连接两个节点的连线称为分枝，根据所处的位置，可能代表某一策略(方案枝)，或代表某一状态(概率枝)，在决策树中，常在被淘汰的策略(方案)的分枝上画上两根平行的短杠，将

北京大学出版社版权所有 禁止转载

该分枝切断，表示不选这一方案。

注意：根节点一定是决策节点，这是因为决策树归根结底是为了选择决策方案。

【例 6.13】某企业为增加销售，拟定开发一个新产品。有两种方案可供选择。

方案一：投资 400 万元，建大车间。建成后，如果销路好，每年获利 75 万元；如果销路差，每年将亏损 10 万元，使用年限 10 年。

方案二：投资 150 万元，建小车间。建成后，如果销路好，每年获利 30 万元；如果销路差，每年将盈利 5 万元，使用年限 10 年。

据市场调查预测，新产品在今后 10 年内，销路好的概率是 0.7，销路差的概率是 0.3。决策哪个方案好？

解：采用决策树法进行决策。

1) 依据题意，画出如图 6.2 所示的决策树图

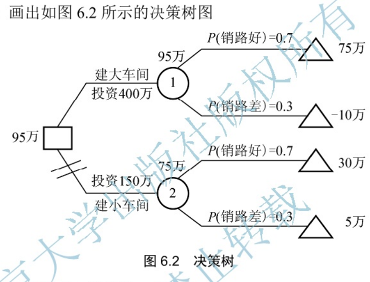


图 6.2 决策树

北京大学出版社版权所有 禁止转载

2) 从右到左，计算各节点期望值

节点 1 的期望值=$\{[0.7\times75+0.3\times(-10)]\times10-400\}$万元$=95$ 万元

节点 2 期望值=$\{[0.7\times30+0.3\times5]\times10-150\}$万元$=75$ 万元

3) 比较节点 1 和 2 的期望值，舍去节点 2，决策结果为选择建大车间的方案

应用决策树来做决策的过程，是从右向左逐步后退进行分析。根据右端的损益值和概率枝的概率，计算出期望值的大小，确定方案的期望结果，然后根据不同方案的期望结果作出选择。方案的舍弃叫做修枝，被舍弃的方案用“≠”的记号来表示，最后的决策点留下一条树枝，即为最优方案。

当所要的决策问题只需进行一次决策就可解决时，叫做单阶段决策问题，例 6.9 就是单阶段决策问题。如果问题比较复杂，而要进行一系列的决策才能解决，就叫做多阶段决策问题，多阶段决策问题采用决策树决策方法比较直观容易。与损益表法相比，决策树法有许多优点，损益表只能表示单阶段决策问题，且要求所有行动方案所面对的自然状态完全一致。

【例 6.14】某企业为了适应市场发展的需要，提出了扩大微波炉生产的两个方案，一个方案是建设大工厂，第二个方案是建设小工厂。建设大工厂需要投资 600 万元，建设小工厂投资 280 万元，都可使用 10 年。估计在此期间，产品销路好的概率是 0.7；销路差的概率是 0.3，建大工厂，销路好可每年赚 200 万元，销路差就要亏 40 万元。建小工厂，销路好每年可赚 80 万元，销路差也可赚 60 万元，如若销路好的市场状态出现，3 年后扩建，扩建需要投资 400 万元，可使用 7 年，每年赢利 190 万元。根据经验知识，后 7 年的市场情

况与前 3 年的市场状态一致，试用决策树法选出合理的决策方案。

解： 图 6.3 是例 6.14 的决策树，计算各点的期望值为

节点 2：{[0.7×200+0.3×(−40)]×10−600(投资)}万元=680 万元

节点 5：{190×7−400}万元=930 万元

节点 6：(80×7)万元=560 万元

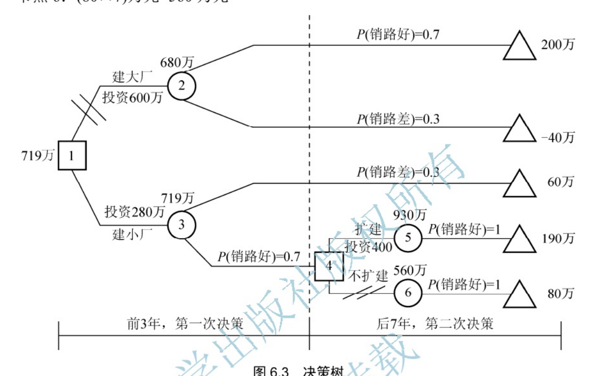


图 6.3 决策树

比较决策节点 4 的方案。由于节点 5(930 万元)与节点 6(560 万元)相比，节点 5 的期望利润值较大，因此应采用扩建的方案，而舍弃不扩建的方案。把节点 5 的 930 万元移到节点 4 来，可计算出节点 3 的期望利润值。

节点 3 的期望利润为[0.7×80×3+0.7×930+0.3×60×(3+7)−280]万元=719 万元。

最后比较决策节点 1 的方案。由于节点 3(719 万元)与节点 2(680 万元)相比，节点 3 的期望利润值较大，因此取节点 3 而舍节点 2。这样，相比之下，建设大工厂的方案不是最优方案，合理的策略应采用前 3 年建小工厂，如销路好，后 7 年进行扩建。

【例 6.15】 从事石油钻探工作的 B 企业与某石油公司签订合同，在一片含油的荒地钻探。B 企业可采用先做地震试验，然后再由试验结果决定是否钻井；或不用地震试验，只凭自己的经验来决定是否钻井。已知做地震试验的费用每次为 0.3 万元，钻井费用为 1 万元。若 B 企业钻井出油后，它可以收入 4 万元；若钻井但不出油，它将无任何收入。其先验概率和条件概率见表 6-14，试就该问题进行贝叶斯决策。

表 6-14 先验概率和条件概率

状 态	先验概率 $P(\theta_i)$	条件概率 $P(H_k/\theta_i)$	
		试验结果好 H_1	试验结果不好 H_2
出油 θ_1	0.55	0.93	0.07
不出油 θ_2	0.45	0.2	0.8

北京大学出版社版权所有 禁止转载

解：这里采用决策树法进行贝叶斯决策。

(1) 首先计算后验概率，见表 6-15。

表 6-15 后验概率计算表

		自然状态和先验概率		
		出油 θ_1	不出油 θ_2	
		$P(\theta_1)$ =0.55	$P(\theta_2)$ =0.45	
条件概率 $P(H_k\|\theta_j)$	H_1	0.93	0.20	
	H_2	0.07	0.80	
乘积概率 $P(H_k\|\theta_j)P(\theta_j)$	H_1	0.511 5	0.09	$P(H_1)=0.60$
	H_2	0.038 5	0.36	$P(H_2)=0.40$
后验概率 $P(\theta_j\|H_k)$	H_1	0.85	0.15	
	H_2	0.10	0.90	

(2) 建立决策树，如图 6.4 所示。

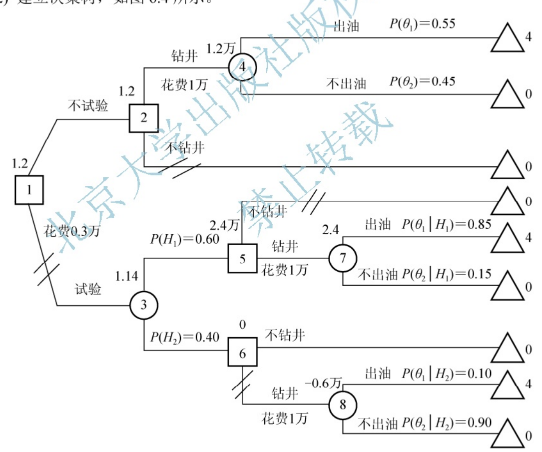


图 6.4 决策树

(3) 最后对该决策树进行分析，计算各点的期望值。

节点 4：(0.55×4+0.45×0−1)万元=1.2 万元

节点 7：(0.85×4+0.15×0−1)万元=2.4 万元

节点 8：(0.10×4+0.90×0−1)万元=−0.6 万元

北京大学出版社版权所有 禁止转载

比较决策节点 2 的方案。由于钻井(1.2 万元)与不钻井方案(0 万元)相比，节点 4 的期望收益值较大，因此决策节点 2 选择钻井的决策方案，把节点 4 的 1.2 万元移到节点 2 来。同理决策节点 5 选择钻井的方案，决策节点 6 选择不钻井的方案。

节点 3 的期望收益为(0.60×2.4+0.40×0−0.3)万元=1.14 万元，最后比较决策节点 1 的方案。由于节点 2(1.2 万元)与节点 3(1.14 万元)相比，节点 2 的期望收益较大，因此取节点 2。

经过上面过程分析，最终选择不试验并钻井的决策方案序列。

6.2.5 风险型决策分析的条件

人们做一项决策，可能成功也可能失败，也就是决策者要冒一定风险。企业经营管理工作遇到的决策问题很多是属于风险型决策问题。风险型决策是指每个备选方案都会遇到几种不同的可能情况，而且已知出现每一种情况的可能性有多大，即发生的概率有多大，因此在依据不同概率所拟定的多个决策方案中，不论选择哪一种方案，都要承担一定的风险。在多数情况下，要获得较高收益的决策，往往要冒较大的风险。对决策者来说，问题不在于敢不敢冒险，而在于能否估计到各种决策方案存在的风险程度，以及在承担风险时所付出的代价与所得收益之间作出慎重的权衡，以便采取行动。

风险型决策方法不同于确定型决策方法，其适用的条件也不同。应用风险型决策方法必须具备以下条件。

(1) 具有决策者期望达到的明确目标。

(2) 存在决策者可以选择的两个以上的可行备选方案。

(3) 存在决策者无法控制的两种以上的自然状态(如气候变化、市场行情和经济发展动向等)。

(4) 不同行动方案在不同自然状态下的收益值或损失值(简称损益值)可以计算出来。

(5) 决策者能估计出不同的自然状态发生的概率。

北京大学出版社版权所有 禁止转载

6.3 不确定型决策

不确定型决策的条件与风险型决策基本相同，只是无法测算各种状态出现的概率。这时的决策主要取决于决策者的经验、智能和思维判断。由于决策者面临哪一种自然状态是完全不确定的，因而决策的结果也是完全不确定的，所以称为不确定型决策。不确定型决策有悲观主义准则、乐观主义准则、乐观系数准则、机会均等准则和后悔值准则。

6.3.1 悲观主义准则

悲观主义准则也称保守主义决策准则或最大最小决策准则。当决策者面临着事物未来的自然状态的发生概率不清楚时，决策者要考虑由于错误的决策而造成重大经济损失。由于自己的经济实力比较脆弱，他在处理问题时就比较谨慎或保守。他从各方案的损益值中找出最坏的可能结果，再从其中选择出最好的可能结果，并把它作为决策结果。用符号 max min 表示这种决策准则，采用这种方法是非常保守的，决策者惟恐决策失误造成较大的经济损失。因此在进行决策分析时，从最不利的客观条件出发来考虑问题，力求损失最小。

具体做法是：在收益矩阵中先从各方案所对应的各种自然状态下找出其收益的最小值，

将它们列于矩阵表的最右列；再从此列中选出数值最大者，并以它对应的策略作为决策者应选的方案。上述计算过程可用公式表示为

$$\max_i \min_j \{u_{ij}\} \to d_k$$

【例 6.16】某公司为了扩大市场，要举行一个展销会。会址打算选择甲、乙、丙三地。获利情况除了与会址有关系外，还与天气有关，天气可区分为晴、普通、多雨三种。采用不同方案在不同天气条件下的公司收益见表 6-16，试用悲观主义准则确定决策方案。

表 6-16　损益表

状态 方案	晴 θ_1	普通 θ_2	多雨 θ_3	最小收益值
甲地 A_1	100	40	30	30
乙地 A_2	70	50	60	50
丙地 A_3	30	25	50	25
最小收益值中的最大收益				50
所选定的决策方案				A_2

解：求解过程参见表 6-16，悲观主义准则确定的决策方案为 A_2。

6.3.2　乐观主义准则

持乐观主义(max max)准则的决策者，对待风险的态度与悲观主义者不同。当他面临情况不明的策略问题时，他宁可冒风险而决不放弃任何一个可获得最好收益的机会，以争取好中求好的乐观主义态度来选择其策略。这种方法的特点是，决策者对决策事件未来前景的估计乐观并有成功的把握。因此愿意以承担风险的代价去获得最大收益。

决策者在分析收益矩阵各策略所对应的各种自然状态时，从中选择出最大收益值列于表的最右列；再从该列数值中选出最大值，以它所对应的策略作为决策方案。上述决策过程可用公式表示为

$$\max_i \max_j \{u_{ij}\} \to d_k$$

例 6.16 采用乐观主义准则确定的决策方案为 A_1，见表 6-17。

表 6-17　损益表

状态 方案	晴 θ_1	普通 θ_2	多雨 θ_3	最大收益值
甲地 A_1	100	40	30	100
乙地 A_2	70	50	60	70
丙地 A_3	30	25	50	50
最大收益值中的最大收益				100
所选定的决策方案				A_1

北京大学出版社版权所有 禁止传播

6.3.3 乐观系数准则

这是介于上述两个准则之间的一个准则，把自然状态好或差的概率变成人为地估计一种可能性，对乐观和悲观出现的可能性估计就是乐观系数。

决策人根据市场预测和经验判断确定一个乐观系数α为主观概率，其值为$0\leqslant\alpha\leqslant1$，则每个方案的估计损益期望值为

$$\alpha\times\text{最大损益值}+(1-\alpha)\times\text{最小损益值}$$

如在例 6.16 取乐观系数α=0.8，则

方案A_1的损益期望值=0.8×100+0.2×30=86

方案A_2的损益期望值=0.8×70+0.2×50=66

方案A_3的损益期望值=0.8×50+0.2×25=45

根据各个方案估算损益期望值的大小，选择最大值 86 为决策方案，故应选方案A_1。

乐观系数准则比较接近实际，但乐观系数的决定非常关键，常带有决策者的主观性。

6.3.4 机会均等准则

假定各个自然状态发生的概率相等，计算各个方案期望损益值，再比较期望损益值，选择最大期望值的方案。也就是说，对各种可能出现的状态“一视同仁”，再按照期望收益最大的原则选择最优方案。

如在例 6.16 中，各个方案的等可能期望值为

方案A_1的损益期望值=(100+40+30)×1/3=56.67

方案A_2的损益期望值=(70+50+60)×1/3=60

方案A_3的损益期望值=(30+25+50)×1/3=35

故以期望值最大的A_2方案为最优方案。

北京大学出版社版权所有 禁止转载

6.3.5 后悔值准则

这是考虑决策机会成本的准则，如果决策者当初并未采用这一方案，而采用了其他方案，必然感到后悔，后悔当初未选择最大收益值的方案。为避免决策的失误造成机会损失而后悔，决策目的是使折中后悔减少到最低程度，以各个方案机会损失大小来判定方案的优劣。

决策过程是在计算出各个方案在各种自然状态下的后悔值以后，从中选择每个方案的最大后悔值，然后从最大后悔值中选取最小者为决策方案。当某一自然状态发生时，即可明确哪个方案是最优的，其收益值是最大的。在所有方案中，将每一种自然状态选出最大收益值定为理想目标值，最大收益值与所采取方案的收益值之差，称为后悔值。找出各个方案的最大后悔值，从中选择后悔值最小的方案作为最佳方案。

如在例 6.16 中，按后悔值准则决策，参见表 6-18，应采用方案A_1或A_2。

表 6-18 损益表

方案 \ 状态	θ_1	θ_2	θ_3	后悔值			最大后悔值
				θ_1	θ_2	θ_3	
A_1	100	40	30	0	10	30	30
A_2	70	50	60	30	0	0	30
A_3	30	25	50	70	25	10	70
最大后悔值中的最小值				30			
所选定的决策方案				A_1 或 A_2			

以上 5 种准则作为不确定型决策优选方案的依据，都带有相当程度的随意性，从例 6.16 中可以看出，不同决策准则的决策的结果是不一样的。因此，在具体管理工作中，决策准则的选择，还要取决于决策者的知识、经验、观念、综合分析判断能力和魄力。

6.4 马尔科夫决策

马尔科夫决策方法是应用马尔科夫链的基本原理和基本方法分析预测随机事件未来发展变化的趋势，也就是通过随机变量的现在状态和动向预测其将来的状态和动向，以便采取相应的对策，马尔科夫决策方法在市场预测和营销决策等方面有广泛的应用。

6.4.1 马尔科夫决策模型

在实际中常常遇到随机运动的系统，系统不断地改变它的状态。马尔科夫模型是描述这类系统的有力工具，它是最简单的一类随机过程，在自然科学和社会科学的各个领域都有重要的作用。

定义 6.1 概率向量：若向量 $\boldsymbol{u}=(u_1\ u_2\ \cdots\ u_n)^{\mathrm{T}}$ 满足 $u_i \geqslant 0(i=1,2,\cdots,n)$ 且 $\sum_{i=1}^{n} u_i = 1$，则称 $\boldsymbol{u}$ 为概率向量。若矩阵 $\left[a_{ij}\right]_{n\times n}$ 各个行向量均为概率向量，称矩阵 $\left[a_{ij}\right]_{n\times n}$ 为随机(概率)矩阵。

三阶方阵 $\boldsymbol{P}=\begin{bmatrix} 0.8 & 0.15 & 0.05 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.2 & 0.7 \end{bmatrix}$ 就是概率矩阵。

定义 6.2 马尔科夫链：设 $X_1, X_2, \cdots, X_n, \cdots$ 为离散的随机变量序列，简记为 $\{X_n\}$，X_n 的所有可能取值的全体称为 $\{X_n\}$ 的状态空间，记为 $E=\{x_1, x_2, x_3, \cdots\}$。若对任意的正整数 n 及任意的 $x_{i_1}, x_{i_2}, x_{i_3}, \cdots, x_{i_n}, x_{i_{n+1}} \in E$，只要

$$P(X_1 = x_{i_1}, X_2 = x_{i_2}, X_3 = x_{i_3}, \cdots, X_n = x_{i_n}) > 0$$

就有

$$P(X_{n+1} = x_{i_{n+1}} \mid X_1 = x_{i_1}, X_2 = x_{i_2}, X_3 = x_{i_3}, \cdots, X_n = x_{i_n}) = P(X_{n+1} = x_{i_{n+1}} \mid X_n = x_{i_n})$$

则称 $\{X_n\}$ 为马尔科夫链，简称马氏链。

北京大学出版社版权所有 禁止转载



定义 6.3 齐次马氏链：$\{X_n\}$为马氏链，若对任意的$x_i, x_j \in E$，总有

$$P(X_{n+1}=x_j | X_n = x_i) = P(X_{m+1}=x_j | X_m = x_i)$$

则称$\{X_n\}$为齐次马氏链。

定义 6.4 转移矩阵：若$\{X_n\}$为齐次马氏链，称$P(X_{n+k}=x_j | X_n = x_i)$为$\{X_n\}$从状态$x_i$到状态$x_j$的$k$步转移概率，记作$p_{ij}(k)$；称以$p_{ij}(k)$（$x_i, x_j \in E$）为元素的矩阵为$\{X_n\}$的$k$步转移矩阵，记作$\boldsymbol{P}(k)$，特别地，将一步转移概率和一步转移矩阵分别记为$p_{ij}$和$\boldsymbol{P}$。显然，一步和多步转移矩阵都是随机矩阵。

定义 6.5 极限矩阵：若矩阵 $\boldsymbol{A}(n)=\begin{bmatrix} a_{11}(n) & a_{12}(n) & \cdots & a_{1m}(n) \\ a_{21}(n) & a_{22}(n) & \cdots & a_{2m}(n) \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}(n) & a_{m2}(n) & \cdots & a_{mm}(n) \end{bmatrix}$ 每一个元素$a_{ij}(n)$（$n=1,2,3,\cdots$）都是数列$\{a_{ij}(n)\}$的项，称矩阵$\boldsymbol{A}(n)$为数列矩阵；若对任意的$i,j=1,2,\cdots,m$都有$\lim\limits_{n\to\infty} a_{ij}(n)=a_{ij}$成立，即每个数列的极限都存在，称矩阵$\boldsymbol{A}(n)$在$n\to\infty$时的极限为$\boldsymbol{A}=[a_{ij}]$，$\boldsymbol{A}$是$\boldsymbol{A}(n)$的极限矩阵。

定理 6.1 齐次马氏链的k步转移矩阵$\boldsymbol{P}(k)=\boldsymbol{P}^k$。

定理 6.2 若齐次马氏链的k步转移矩阵$\boldsymbol{P}(k)$在$k\to\infty$的极限矩阵存在，记为$\lim\boldsymbol{P}$，则$\lim\boldsymbol{P}$是一个随机矩阵。该随机矩阵也是一个转移矩阵，且对任意的自然数n，都有$(\lim\boldsymbol{P})^n=\lim\boldsymbol{P}$。

若齐次马氏链的k步转移矩阵$\boldsymbol{P}(k)$的极限矩阵存在，应用定理 6.2，则随着系统的不断演化，最终系统状态之间的转移概率保持不变，系统体现出统计规律性的特征，就演化为一个稳定的系统。为了研究问题方便起见，在实际问题中经常考虑有限个状态的齐次马氏链，它们的k步转移矩阵的极限矩阵总存在。

【例 6.17】 在我市销售的鲜牛奶主要由 3 个厂家提供。去年 12 月份对 2 000 名消费者进行了调查，其中买厂家 1 的消费者有 800 人，买厂家 2 和厂家 3 的消费者均为 600 人。调查还显示有的消费者在今年 1 月份继续购买原厂家的牛奶，而有的消费者转向购买其他厂家的产品，具体调查结果见表 6-19。

表 6-19 牛奶市场消费调查表 单位：人数

		1 月份			
		厂家 1	厂家 2	厂家 3	合 计
12 月份	厂家 1	320	240	240	800
	厂家 2	360	180	60	600
	厂家 3	360	120	120	600

(1) 试写出我市牛奶消费者变化情况的一步转移概率矩阵；

(2) 若某消费者在去年 12 月份购买厂家 1 的牛奶，则他在今年 2 月份购买厂家 3 的概率是多大？

北京大学出版社版权所有 禁止转载

解：(1) 根据调查统计表，可得状态转移矩阵为

$$\boldsymbol{P}=\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}=\begin{bmatrix} 320/800 & 240/800 & 240/800 \\ 360/600 & 180/600 & 60/600 \\ 360/600 & 120/600 & 120/600 \end{bmatrix}=\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.6 & 0.3 & 0.1 \\ 0.6 & 0.2 & 0.2 \end{bmatrix}$$

(2) 消费者在去年 12 月份购买厂家 1 的牛奶，则他在今年 2 月份购买厂家 3 的概率可由转移矩阵计算出来，即求消费者由厂家 1 经两次转移到厂家 3 的概率。

根据定理 6.1，先求两步转移矩阵为

$$\boldsymbol{P}(2)=\boldsymbol{P}^2=\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.6 & 0.3 & 0.1 \\ 0.6 & 0.2 & 0.2 \end{bmatrix}\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.6 & 0.3 & 0.1 \\ 0.6 & 0.2 & 0.2 \end{bmatrix}$$

所求概率为 $\boldsymbol{P}(2)$ 中第一行第三列中元素值，即

$$\begin{bmatrix} 0.4 & 0.3 & 0.3 \end{bmatrix}\times\begin{bmatrix} 0.3 & 0.1 & 0.2 \end{bmatrix}^{\mathrm{T}}=0.4\times0.3+0.3\times0.1+0.3\times0.2=0.21。$$

6.4.2 马尔科夫决策模型的应用

在经济等领域中，有许多系统具有马尔科夫链的特征，因此可以运用其方法原理加以分析研究，作出科学的预测决策。

【例 6.18】 设某一地段有两家酒店，每天到这两家酒店用餐的顾客总数是固定的，不妨设为 200 人，通过统计发现，今日在 A 酒店用餐的顾客，明天还在 A 酒店用餐的概率是 0.8，而明天转到 B 酒店用餐的概率是 0.2；同理，今日在 B 酒店用餐的顾客，明天还在 B 酒店用餐的概率是 0.7，而明天转到 A 酒店用餐的概率是 0.3，那么，经过一段时间后，这 200 个顾客在这两家酒店用餐的分布情况如何？

解：对一个顾客来说，他只有两种选择，要么到 A 酒店，要么就到 B 酒店。用概率论的语言来描述，这个系统只有两种状态 A 和 B，顾客可以在这两个状态中相互转移。虽然每次的转移是随机的、偶然的，但在大量和长期的转移过程中还是有一定统计规律，其转移具有以下规律。

(1) A→A 的概率是 0.8，A→B 的概率是 0.2；B→B 的概率是 0.7，B→A 的概率是 0.3，数学上可用一个称之为“转移概率矩阵”来描述，见表 6-20。

表 6-20 转移概率

现状态	下步状态	
	A	B
A	0.8	0.2
B	0.3	0.7

相应的转移矩阵为 $\boldsymbol{P}=\begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$。

(2) 上述的转移过程满足：“若已知现在所处的状态，则将来的转移情况与过去所处的状态无关”，显然该转移过程是一个齐次马氏链。

现在假设开始到 A、B 酒店用餐的顾客各有 100 名，按平均意义，一天后原来在 A 店的 100 名顾客中有 $100\times0.8=80$ 仍然到 A 店用餐，而有 $100\times0.2=20$ 名转移到 B 店用餐；

北京大学出版社版权所有 禁止转载

同理，原来在 B 店的 100 名顾客中有$100\times 0.3=30$ 名转移到 A 店用餐，而仍然有$100\times 0.7=70$名在 B 店用餐，若$\boldsymbol{M}_n$表示第n天后到 A、B 酒店用餐的人数，那么第一天的用餐人数矩阵是$\boldsymbol{M}_0=(100\quad 100)$，用矩阵表示上述转移情况就得

$$\boldsymbol{M}_1=(100\quad 100)\begin{pmatrix}0.8 & 0.2\\ 0.3 & 0.7\end{pmatrix}=\boldsymbol{M}_0\boldsymbol{P}$$

同理$\boldsymbol{M}_2=\boldsymbol{M}_1\boldsymbol{P}\quad \boldsymbol{M}_3=\boldsymbol{M}_2\boldsymbol{P}\cdots \boldsymbol{M}_n=\boldsymbol{M}_{n-1}\boldsymbol{P}$

一般地有$\boldsymbol{M}_n 0=\boldsymbol{M}_{n-1}\boldsymbol{P}=\boldsymbol{M}_{n-2}\boldsymbol{P}\boldsymbol{P}=\boldsymbol{M}_{n-2}\boldsymbol{P}^2=\boldsymbol{L}=\boldsymbol{M}_0\boldsymbol{P}^n$，依次发展下去。因为有限个状态的齐次马氏链的极限矩阵总存在，则有$\lim\limits_{n\to\infty}\boldsymbol{M}_n=\lim\limits_{n\to\infty}\boldsymbol{M}_{n-1}\boldsymbol{P}$，也就有$[x\quad y]=[x\quad y]\boldsymbol{P}$，根据两个矩阵相等的充要条件是它们的对应元素相等，得

$$\begin{cases}x=0.8x+0.3y\\ y=0.2x+0.7y\end{cases}$$

因两个方程同解，加上总人数是固定的，可得下面方程

$$\begin{cases}x+y=200\\ y=0.8x+0.3y\end{cases}$$

解得$x=120, y=80$，表明如果不改变两家酒店的经营方式，则一般 A、B 酒店每天有 120 个和 80 个顾客。

需要注意的是，虽然每个人到各个酒店用餐是随机的，但随着时间的推移，到各个酒店用餐的人数是固定的，体现出一定的统计规律性，这是酒店管理者需要了解的信息，以便进一步提高管理和服务水平。

【例 6.19】银行为了对不良债务的变化趋势进行分析和预测，经常按以下方式对贷款情况进行划分：

N_1——逾期贷款，拖延支付本金利息达 0～60 天；

N_2——怀疑贷款，拖延支付本金利息达 61～180 天；

N_3——呆滞贷款，拖延支付本金利息达 181～360 天；

N_4——呆账贷款，拖延支付本金利息达 360 天以上；

N_5——付清本金利息。

若某银行当前贷款总额 470 万元，其中属于逾期贷款N_1=200 万元；属于怀疑贷款N_2=150 万元；属于呆滞贷款N_3=120 万元，据隔月账面变化情况分析，近似得状态之间的转移矩阵为

$$\boldsymbol{P}=\begin{bmatrix}0.30 & 0.30 & 0.00 & 0.00 & 0.40\\ 0.15 & 0.25 & 0.30 & 0.00 & 0.30\\ 0.10 & 0.10 & 0.30 & 0.35 & 0.15\\ 0.00 & 0.00 & 0.00 & 1.00 & 0.00\\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00\end{bmatrix}$$

其中p_{ij}表示当前处于N_i状态的贷款，1 月后处于N_j的概率，当$i>j$时，如p_{21}表示当前贷款处在已欠 61～180 天的N_2状态，而在本月中贷款者归还了部分本金利息，贷款转为状态N_1的转移概率。

令X_t表示第t月的贷款分布，$E=\{N_1,N_2,N_3,N_4,N_5\}$表示状态空间，据上述信息，对

北京大学出版社版权所有 禁止转载

不良贷款的变化趋势进行研究，需求 k 步转移矩阵 $\boldsymbol{P}(k)$ 的极限矩阵，应用矩阵特征值理论，得 $\boldsymbol{P}=\boldsymbol{M}^{-1}\boldsymbol{D}\boldsymbol{M}$，其中

$$\boldsymbol{D}=\begin{bmatrix}0.15 & 0.00 & 0.00 & 0.00 & 0.00\\0.00 & 0.60 & 0.00 & 0.00 & 0.00\\0.00 & 0.00 & 0.10 & 0.00 & 0.00\\0.00 & 0.00 & 0.00 & 1.00 & 0.00\\0.00 & 0.00 & 0.00 & 0.00 & 1.00\end{bmatrix}\quad \boldsymbol{M}=\begin{bmatrix}-0.78 & -2.33 & 4.67 & -1.92 & 0.37\\0.52 & 0.63 & 0.63 & -0.55 & -1.23\\0.00 & -2.91 & 4.37 & -1.70 & 0.24\\0.00 & 0.00 & 0.00 & 1.17 & 0.00\\0.00 & 0.00 & 0.00 & 0.00 & 1.61\end{bmatrix}$$

所以

$$\lim\boldsymbol{P}=\begin{bmatrix}0.00 & 0.00 & 0.00 & 0.10 & 0.90\\0.00 & 0.00 & 0.00 & 0.24 & 0.76\\0.00 & 0.00 & 0.00 & 0.55 & 0.45\\0.00 & 0.00 & 0.00 & 1.00 & 0.00\\0.00 & 0.00 & 0.00 & 0.00 & 1.00\end{bmatrix}$$

$$\begin{bmatrix}200 & 150 & 120 & 0 & 0\end{bmatrix}\begin{bmatrix}0.00 & 0.00 & 0.00 & 0.10 & 0.90\\0.00 & 0.00 & 0.00 & 0.24 & 0.76\\0.00 & 0.00 & 0.00 & 0.55 & 0.45\\0.00 & 0.00 & 0.00 & 1.00 & 0.00\\0.00 & 0.00 & 0.00 & 0.00 & 1.00\end{bmatrix}$$

$$=\begin{bmatrix}0 & 0 & 0 & 122.46 & 347.54\end{bmatrix}$$

通过上述计算可知，最终可能有 122.46 万元成为呆账，只有 347.54 万元可能收回；也可以得到，逾期贷款中，约 10%可能成为呆账，90%可能收回；怀疑贷款中，约 24%可能成为呆账，76%可能收回；呆滞贷款中，约 55%可能成为呆账，45%可能收回(见矩阵 $\lim\boldsymbol{P}$)。根据可回收的贷款预测数，便于合理地设定贷款利率。

【例 6.20】某出租汽车公司为宏观管理车辆，需对本公司的汽车流向进行预测。显然，出租汽车的流向是一个随机过程。设某市共有 4 个区，出租汽车在这 4 个区运营。假设出租汽车平均运营一次用 1h，在 n 时刻某车在 i 区接业务在 $n+1$ 时刻到达 j 区完成业务下车。再设 X_n 表示 n 时刻出租车所在的区域，显然 $\{X_n\}$ 是一个马氏链，因为汽车在 $n+1$ 时刻到达 j 区只与 n 时刻在 i 区接业务有关，接着设出租车从 i 区到 j 区的转移概率 $p_{ij}=P\{X_{n+1}=j|X_{n=i}=i\}$ 与初始时刻 n 无关，则 $\{X_n\}$ 是一个齐次马氏链，且状态空间 $E=\{1,2,3,4\}$。

经统计，可进一步得出转移矩阵为

$$\boldsymbol{P}=\begin{bmatrix}0.4 & 0.2 & 0.3 & 0.1\\0.3 & 0.3 & 0.3 & 0.1\\0.2 & 0.2 & 0.4 & 0.2\\0.4 & 0.1 & 0.2 & 0.3\end{bmatrix}$$

求出 k 步转移矩阵 $\boldsymbol{P}(k)$ 的极限矩阵为

北京大学出版社版权所有 禁止转载

$$\lim \boldsymbol{P}=\begin{bmatrix}0.317 & 0.204 & 0.315 & 0.164\\0.317 & 0.204 & 0.315 & 0.164\\0.317 & 0.204 & 0.315 & 0.164\\0.317 & 0.204 & 0.315 & 0.164\end{bmatrix}$$

因为该极限矩阵的各行元素相同，且它们的和等于 1，称它们为马氏链$\{X_n\}$的极限分布，记极限分布为$\Pi=(0.307, 0.204, 0.315, 0.164)$，根据该极限分布可预测出各区运营车辆的情况，以便对车辆进行调控，便于交通管理及停车场的安排。

【例 6.21】某鞋厂生产高、中、低档三种类型的皮鞋，每月统计一次销售变化情况。令X_n表示客户定购的鞋型，则它是一个随机变量，其状态空间为$E=\{1,2,3\}$，其中 1、2、3 分别表示高、中、低档皮鞋。经统计有一步转移矩阵为

$$\boldsymbol{P}=[p_{ij}]=\begin{bmatrix}0.5 & 0.3 & 0.2\\0.4 & 0.4 & 0.2\\0.3 & 0.3 & 0.4\end{bmatrix}$$

其中$p_{ij}=P\{X_{n+1}=j|X_{n=i}=i\}$表示上月定购$i$型鞋的客户下月定购$j$型鞋的概率$(i,j=1,2,3)$。

假设客户转移概率不变，则$\{X_n\}$是一个齐次马氏链，求出其k步转移矩阵$\boldsymbol{P}(k)$的极限矩阵为

$$\lim \boldsymbol{P}=\begin{bmatrix}0.42 & 0.33 & 0.25\\0.42 & 0.33 & 0.25\\0.25 & 0.33 & 0.25\end{bmatrix}$$

北京大学出版社版权所有 禁止转载

马氏链的极限分布为$\Pi=(0.42, 0.33, 0.25)$，也就是客户对高、中、低档皮鞋的稳定购买率，对企业制订生产计划是极其重要的。

【例 6.22】某商店经营一种易腐食品，出售一个单位可获利 5 元。若当天售不出去，则每单位损失 3 元。该店经理统计了连续 40 天的需求情况(不是实际销售量)，统计数据如下：

3，3，4，2，2，4，2，3，4，4，4，3，2，4，2，3，3，4，2，2
4，3，4，3，2，3，4，2，3，2，2，3，4，2，4，4，3，2，3，3

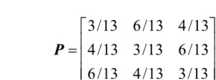
经理想应用马氏链来预测需求量，确定明日进货量。(1)已知当天需求量为 3 个单位，明日应进货多少个单位？(2)若不知当天需求量，明日应进货多少个单位？

解：设状态N_1=需求量为 2 个单位，状态N_2=需求量为 3 个单位，状态N_3=需求量为 4 个单位，则据统计结果可知状态转移的情况见表 6-21。

表 6-21 三种状态转移情况表

	第二天状态 N_1	第二天状态 N_2	第二天状态 N_3
第一天状态 N_1	3	6	4
第一天状态 N_2	4	3	6
第一天状态 N_3	6	4	3

因此该马氏链的状态转移概率矩阵为



$$
\boldsymbol{P}=\begin{bmatrix} 3/13 & 6/13 & 4/13 \\ 4/13 & 3/13 & 6/13 \\ 6/13 & 4/13 & 3/13 \end{bmatrix}
$$



其极限矩阵为

$$\lim \boldsymbol{P}=\begin{bmatrix}1/3 & 1/3 & 1/3\\1/3 & 1/3 & 1/3\\1/3 & 1/3 & 1/3\end{bmatrix}$$

设明天的需求量为q个单位(显然q为随机变量)，明天的进货量为s个单位(s为决策变量)，则相应的利润为

$$L(s,q)=\begin{cases}5s & q\geqslant s\\8q-3s & q<s\end{cases}$$

(1) 若当天需求量为 3 个单位，属于状态N_2，一步转移概率见矩阵P，则期望利润为

$$f(s)=E(L(s,q))=\frac{4}{13}L(s,2)+\frac{3}{13}L(s,3)+\frac{6}{13}L(s,4)=\begin{cases}5s & s\leqslant 2\\ \dfrac{64}{13}+\dfrac{33}{13}s & 2<s\leqslant 3\\ \dfrac{136}{13}+\dfrac{9}{13}s & 3<s\leqslant 4\\ \dfrac{328}{13}-3s & s>4\end{cases}$$

由最大期望值准则得

$$\max_s f(s)=f(4)=172/13$$

因此，若当天的需求量为 3 个单位，则明天应进货 4 个单位。

(2) 若不知当天需求量，采用马氏链的极限分布概率计算，则期望利润为

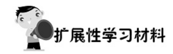
$$f(s)=E(L(s,q))=\frac{1}{3}L(s,2)+\frac{1}{3}L(s,3)+\frac{1}{3}L(s,4)=\begin{cases}5s & s\leqslant 2\\ \dfrac{16}{3}+\dfrac{7}{3}s & 2<s\leqslant 3\\ \dfrac{40}{3}-\dfrac{1}{3}s & 3<s\leqslant 4\\ 24-3s & s>4\end{cases}$$

由最大期望值准则得

$$\max_s f(s)=f(3)=37/3$$

因此，若不知当天需求量，则明天应进货 3 个单位。

马尔科夫是一类经常遇到的模型，特别在随机服务系统和可靠性系统中，马尔科夫模型有着不可替代的作用。



多目标决策

多目标决策是数学规划的一个分支，研究多于一个目标函数在给定区域上的最优化，通常记为 VMP。在很多实际问题中，例如经济、管理、军事、科学和工程设计等领域，衡量一个方案的好坏往往难以用一个指标来判断，而需要用多个目标来比较，而这些目标有时不甚协调，甚至是矛盾的，因此有许多学者致力于这方面的研究。1896 年法国经济学家 V. 帕雷托最早研究不可比较目标的优化问题，之后 J. 冯·诺伊

北京大学出版社版权所有 禁止转载

曼、H. W. 库恩、A. W. 塔克尔、A. M. 日夫里翁等数学家做了深入的探讨，但是尚未有一个完全令人满意的定义。求解多目标规划的方法大体上有以下两种：一种是化多为少的方法，即把多目标化为比较容易求解的单目标或双目标，如主要目标法、线性加权法、理想点法等；另一种叫分层序列法，即把目标按其重要性给出一个序列，每次都在前一目标最优解集内求下一个目标最优解，直到求出共同的最优解。多目标决策除以上方法外还可以应用数据包络分析和层次分析法。层次分析法是由美国运筹学家 Satti 于 20 世纪 70 年代提出的，这是一种定性与定量相结合的多目标决策与分析方法，对于目标结构复杂且缺乏必要的数据的情况更为实用。

本 章 小 结

本章首先从实例引入决策的基本要素——状态空间、策略空间和损益表，给出了决策方法的分类；其次就最常用的风险型决策展开了详细的讨论，阐述了期望收益值等决策准则，探讨了决策树法和贝叶斯决策方法，然后就不确定型决策问题给出了相关的决策准则；最后介绍了马尔科夫决策模型及其决策方法的特点。

北京大学出版社版权所有
禁止转载



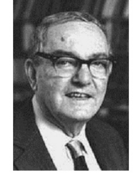
关键术语(中英文)

状态空间(State Space)
决策空间(Decision Space)
损益值(Pay-off Value)
决策准则(Decision Criterion)
确定型决策(Deterministic Type Decision)
风险型决策(Risk Type Decision)
不确定决策(Nondeterministic Type Decision)
期望值准则(Expected Value Criterion)
决策树(Decision Tree)
贝叶斯决策(Bayesian Decision)
EPPI(Expected Profit of Perfect Information)
先验分析(Prior Analysis)
EVPI(Expected Value of Perfect Information)
后验分析(Postprior Analysis)
悲观主义准则(Pessimistic Criterion)
乐观主义准则(Optimistic Criterion)
乐观系数准则(Optimistic Coefficient Criterion)
后悔值准则(Regret Value Criterion)
机会均等准则(Equally Liability Criterion)
马尔科夫决策(Markov Decision)



知识链接

管理决策学派——西蒙



赫伯特·西蒙(Herbert A.Simon，1916—2001)1916 年出生于美国威斯康辛州密歇根湖畔的密尔沃基，父亲是位工程师，母亲是一位很有造诣的钢琴家。西蒙从小就很聪明好学，在密尔沃基的公立学校上学时跳了两级，因此在芝加哥大学注册入学时才 17 岁。1936 年获得芝加哥大学的学士学位，之后从事了几年编辑和行政工作。1943 年获得芝加哥大学政治学博士学位。

西蒙是我们这个时代一个多才多艺的人，他的博士学位是政治学，他的诺贝尔奖是经济学，他在计算机科学、心理学和哲学等领域也有突出的贡献。他多才多艺，兴趣广泛，会画画，会弹钢琴，既爱爬山、旅行，又爱学习各种外国语，能流利地

说多种外语。作为科学家，他涉足的领域之多，成果之丰，影响之深远，令人叹为观止。他和纽厄尔同获图灵奖，是因为他们在创立和发展人工智能方面的杰出贡献，当然是计算机科学家。但是西蒙在 1978 年更荣获诺贝尔经济学奖，不言而喻，是世界一流的大经济学家。1986 年他又因为在行为科学上的出色贡献而荣获美国全国科学奖章(National Medal of Science)。



习题 6

6.1 填空题

(1) 向量 $\boldsymbol{u}=(u_1,u_2,\cdots,u_n)$ 称为概率向量，若 $u_i\geqslant 0(i=1,2,\cdots,n)$，且__________。

(2) 折中决策准则(折中系数 α)假定每个方案中最大收益的概率为__________，最小收益的概率为__________。

(3) 某商店在夏天炎热程度未知情况下，决定是否大批量购进空调设备，这是在__________条件下的决策问题。如果预计天气非常炎热的概率为 40%，这是__________条件下的决策问题。

(4) 如果有两个以上的决策自然条件，但决策人无法估计各自然状态出现的概率，那么这种决策类型称为__________型决策。

(5) 决策方法中，可用于解决多阶段决策问题的是__________。

(6) 设某种产品市场占有率为 $\boldsymbol{T}$，其转换概率矩阵为 $\boldsymbol{P}$，则下一时刻的市场占有率为__________。

6.2 判断题

(1) 属于解决确定型决策的有线性规划、动态规划、盈亏分析和决策树等方法。( )

(2) 属于解决风险型决策问题的基本准则有最大可能准则、渴望水平准则、机会均等准则和期望收益最大准则。 ( )

(3) 在不确定的条件下进行决策，必须：①可拟订出两个以上的可行方案；②可以预测或估计出不同的可行方案在不同的自然状态下的收益值；③确定各种自然状态可能出现的概率值。 ( )

(4) 矩阵 $\begin{bmatrix}0.6 & 0.4\\0.6 & 0.4\end{bmatrix}$、$\begin{bmatrix}0.5 & 0.5\\0.5 & 0.5\end{bmatrix}$、$\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$ 都是概率矩阵。 ( )

(5) 使用马尔科夫决策方法的步骤之一是建立状态转移矩阵。 ( )

6.3 某公司正考虑为开发一种新型产品提供资金。可供选择的方案有 3 个，前景有成功、部分成功与失败。成功的概率为 0.35，部分成功的概率为 0.45，失败的概率为 0.20，其利润见表 6-22，要求：(1)确定最优决策方案；(2)若公司花费 0.2 万元进行一项调查，可以得到关于三个方案前景的更可靠的信息，该公司应进行这项调查吗？

表 6-22 利润 单位：万元

方案 \ 自然状态	成 功	部分成功	失 败
方案 1	20	3	−18
方案 2	15	1	−10
方案 3	10	0	−2

6.4 某服装厂设计了一款新式服装准备推向市场。如直接大批量生产与销售，主观估

计成功与失败的概率各为 0.5，其分别获利为 300 万元与−250 万元，如取消生产销售计划，则损失设计与准备费用 40 万元。根据以上数据，画出决策树按期望收益最大准则确定最优决策。

6.5 某公司有 50 000 元流动资金，如用于某项开发研究估计成功率为 96%，成功时一年可获利 12%，一旦失败，有损失全部资金的危险。如果将资金放到银行里，则可稳得年利 6%。为保证合理投资，公司求助于咨询公司，咨询费用为 500 元，但咨询意见仅供参考，过去咨询公司 200 例意见的实施结果情况见表 6-23，试用决策树分析：

(1) 该公司是否值得求助于咨询公司？

(2) 该公司流动资金应如何合理使用？

表 6-23　咨询公司 200 例意见的实施结果情况　　单位：元

实施结果 / 咨询意见	成　功	失　败	合　计
可以投资	154	2	156
不宜投资	38	6	44
合　计	192	8	200

6.6 国外某石油公司，为了从能源危机中获取巨额利润，试图从油母叶岩中提取石油制品这一事件进行决策。从目前国际形势看，石油价格可能有 4 种变动状态：(1)价格下降，低于现价(低价)；(2)价格上涨，高于现价(高价)；(3)现价不变；(4)出现“禁运”，价格猛涨(禁运)。根据上述情况分析，公司制定在未来 10 年时间内开发油母叶岩的 3 种策略：(1)研究，集中力量研究该油的炼油过程，以降低成本；(2)边研究、边开发，两者结合，但开发人仍按现有工艺进行，会有亏损；(3)应急开发计划，按现有工艺尽快开发，风险更大。各方案的损益值见表 6-24，试求：

(1) 分别按不确定型决策的 5 种方法作出方案选择；

(2) 以你自己的个人价值观来进行决策，你会选取哪一种方案，为什么？

(3) 你如何看待决策中的风险性问题？

表 6-24　各方案的损益值　　单位：百万元

状　态 / 方　案	A 研究	B 边研究、边开发	C 应急开发
低价	−50	−150	−500
现价	0	−50	−200
高价	50	100	0
禁运	55	150	500
低价	−50	−150	−500

6.7 某小镇有两种报纸(A 和 B)。居民人人读报，但只读一种。为了竞争，现两种报纸都拟改版。根据资料估计，改版后，原读报纸 A 的读者仍喜欢读报纸 A 的占 80%，改读报纸 B 的占 20%；原读报纸 B 的读者仍喜欢读报纸 B 的占 70%，改读报纸 A 的占 30%。试问经若干年后，状态达到平衡时，两种报纸的市场占有率各为多少？

6.8 某商场对购买 A、B、C 三种型号的化妆品的顾客作抽样调查：原买 A 型化妆品仍

北京大学出版社版权所有 禁止转载

然继续保持使用 A 型化妆品的人占 80%，改用 B 型化妆品的人占 10%，改用 C 型化妆品的人占 10%。原用 B 型化妆品仍然继续使用 B 型化妆品的人占 90%，改用 C 型化妆品的人占 3%，改用 A 型化妆品的人占 7%。原用 C 型化妆品仍然继续使用 C 型化妆品的人占 70%，改用 B 型化妆品的人占 20%，改用 A 型化妆品的人占 10%。试问：经过一段时间后，处于平衡状态时，这三种型号化妆品的市场占有率分别为多少？



实际操作训练

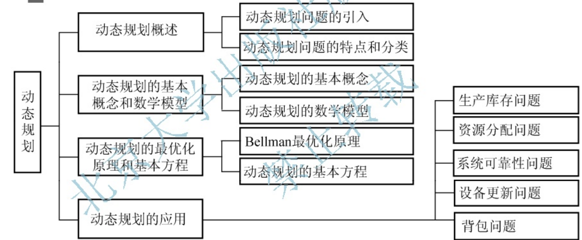
某物流企业承包一建筑工程队的施工设备管理工作，该建筑工程队正在承担一座桥梁的施工任务。由于施工地区夏季多雨，需停工 3 个月。在停工期间可将施工机械搬走或留在原处。如搬走，需搬运费 1 800 元。如留原处，一种方案是花 500 元筑一护堤，防止河水上涨发生高水位的侵袭。若不筑护堤，发生高水位侵袭时将损失 10 000 元；如下暴雨发生洪水时，则不管是否筑护堤，施工机械留在原处都将受到 60 000 元的损失。根据历史资料，该地区夏季高水位的发生率是 25%，洪水的发生率是 2%。试用决策树法分析该物流企业要不要把施工机械搬走？若不搬，则要不要筑护堤？

北京大学出版社版权所有
禁止转载



第7章　动态规划

【本章知识架构】



北京大学出版社版权所有
禁止转载

【本章教学目标与要求】

- 熟练掌握动态规划的基本要素——阶段变量、状态变量、决策变量、状态转移方程、指标函数等。
- 理解动态规划的基本思想，能够建立动态规划的基本方程。
- 熟练应用动态规划的逆序解法去解决一些典型的实际问题。



运储物流公司的设备更新

物流业是资金密集、劳动密集、技术密集的行业，只有加大技术创新力度，加快设备升级改造和信息系统建设，才能提升服务能力，形成差异化竞争优势。物流设备更新有三方面的因素。

(1) 物流设备的自然寿命。它是指物流设备从投入使用开始，直到因为在使用过程中发生物质磨损而不能继续使用、报废为止所经历的时间。

(2) 物流设备的技术寿命。它是指物流设备在市场上维持其价值的时期，是指从物流设备开始使用到因技术落后而被淘汰所延续的时间。

(3) 物流设备的经济寿命。它是指物流设备从投入使用开始，到因继续使用不经济而被更新所经历的时间。

在日益激烈的市场竞争中，运储物流企业运输设备的更新不仅在短期内直接影响到企业的综合竞争能力，而且对企业的长远发展有深远的影响。运储物流公司为了提高设备的利用效率，开始着手制订今后 5 年内的设备更新计划，以使 5 年内的设备使用效益最大化。

资料来源：http:// www.docin.com.

北京大学出版社版权所有 禁止转载

动态规划是运筹学一个重要的分支，在现代企业管理决策中有着广泛的应用。动态规划是20世纪50年代由美国数学家贝尔曼(Richard Bellman)等人建立和发展起来的求解一类多阶段决策问题的优化方法。动态规划的思想是多种算法的基础，比较著名的应用有：最短路径问题、背包问题、网络流优化等。

7.1　动态规划概述

动态规划是解决一类多阶段决策问题的优化方法，提供了分析问题的一种途径和模式。所谓多阶段决策问题，是指一类活动过程，它可按时间或空间把问题分为若干个相互联系的阶段。在每一阶段都要作出选择(决策)，该决策不仅仅决定这一阶段的效益，而且决定下一阶段的初始状态，从而决定整个过程的走向(动态含义)。当每一阶段的决策确定之后，就得到一个决策序列，称为策略。这类问题目的是求一个策略，使各个阶段的效益总和达到最优。如果一个问题可将其过程划分为若干个相互联系的阶段问题，且它的每一阶段都需进行决策，一般可用动态规划方法进行求解。

7.1.1　动态规划问题的引入

下面从实际问题中引出动态规划。

【例 7.1】 一个工厂生产某种产品，需要制订第一季度的生产计划。已知 1—4 月生产成本和产品需求量的变化情况见表 7-1。为了调节生产和需求，工厂设有一个产品仓库，库容量 $H=10$，已知期初库存量为 1，要求期末(4 月底)库存量为 0，每个月生产的产品在月底入库，并根据当月需求发货。求能满足各月的需求，并使总生产成本最低的各月生产量。

表 7-1　成本需求表

月份(k)	1	2	3	4
生产成本(c_k)	13	9	15	18
需求量(d_k)	0	8	5	6

问题是求在满足需求的情况下生产产品，使得生产成本最小。由于每月的生产成本不同，因而每月生产的多少都影响到总成本，所以每月生产量就是管理者要做的决策，某月的生产数量决策又影响其后续月份的生产决策，最终各月生产决策构成一个决策序列。这是一个多阶段决策问题。

【例 7.2】一艘货轮在 A 港装货后驶往 F 港，中途须靠港加油、淡水 3 次，从 A 港到 F 港全部可能的航行路线及两港之间距离如图 7.1 所示，F 港有 3 个码头，试求最合理的停靠的码头及航线，使总路程最短。

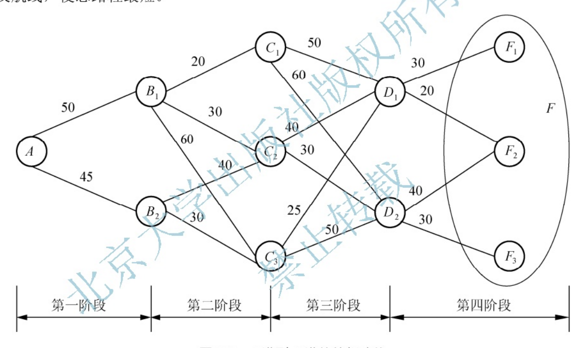


图 7.1　A 港到 F 港的航行路线

北京大学出版社版权所有
禁止转载

问题是求 A 港到 F 港最短路程，但从 A 港到 F 港不能直接到达，中途须靠港加油、淡水 3 次，即必须经过 B_1、B_2，C_1、C_2、C_3，D_1、D_2 中的一个港口加油和加水。这样问题可看成一个多阶段决策问题，第一阶段考虑到 B_1 还是 B_2；第二阶段到 C_1、C_2 还是 C_3？第三阶段考虑到 D_1 还是 D_2；第四阶段考虑到 F_1、F_2 还是 F_3；以此类推，最后获得从 A 港到 F 港的一条路径。显然这样的路径有很多，问题是选择一条最优路径，使得总的路程最短，这是一个典型的动态规划问题。

【例 7.3】物流运作的最短路问题。求从生产厂 Q 到销售商 T 选择哪条路线，有关数据如图 7.2 所示，试求如何选择路线，使得总运费最低？

问题中虽然要求的是运费最低，但实际上还是路程最短问题。根据常识，运费与路程的长短是成正比的。该问题可以按空间活动顺序划分阶段，这是一个多阶段决策问题，它可分为 4 个阶段：第一阶段，从 Q(制造厂)到 A(出口港)；第二阶段，从 A(出口港)到 B(进口港)；第三阶段，从 B(进口港)到 C(城市)；第四阶段，从 C(城市)到 T(销售商)，每个阶段选取的路线不同，对应从 Q 到 T 就有一系列不同的运输路线。该问题考虑的是从生产厂 Q 到

某公司 T 选择哪条路线，使总运费最低(路程最短)？

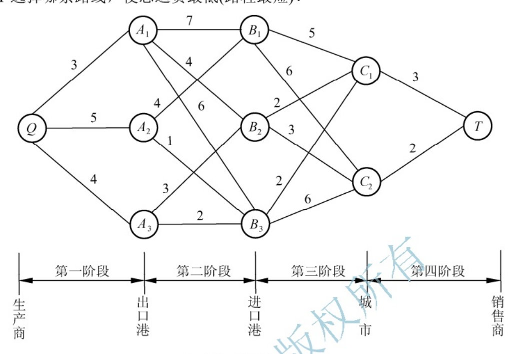


图 7.2 物流运输网络图

北京大学出版社版权所有 禁止转载

从上面实例可知，多阶段决策问题可将其过程划分为若干个相互联系的阶段，且它的每一阶段都需进行决策，该决策既依赖当前的决策条件，又影响以后的决策环境，当每个阶段的决策确定后，就构成了一个决策序列，因而就形成了一个过程完整的活动路线。

一般地，一个 n 阶段的动态规划问题可以描述为图 7.3 所示的模式，其中状态 s_k 表示第 k 个阶段开始时的已知决策条件；u_k 表示在这种决策环境下第 k 个阶段所采取的决策，选择这个方案获得的收益为 $v(u_k)$；该决策也决定了第 $k+1$ 个阶段的状态 s_{k+1}，状态 s_{k+1} 确定依赖于状态之间的关系方程 T_k，$k=1,2,\cdots,n$。

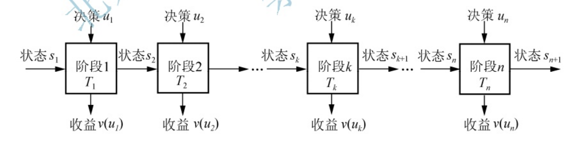


图 7.3 n 阶段的动态规划问题

【例 7.4】 (资源分配问题)设有 6 万元资金用于 4 个工厂的扩建，已知每个工厂的投资效益(万吨)同投资额的大小有关，见表 7-2(表内数据是效益 $g_i(u_j)$，即工厂 i 投资额为 u_j 所能获得的产出)，应如何确定对这 4 个工厂的投资额，使总投资效益最大？

表 7-2 投资收益表　　单位：万吨

投资额(j) 工厂(i)	0	1	2	3	4	5	6
1	0	20	42	60	75	85	90

续表

工厂(i) \ 投资额(j)	0	1	2	3	4	5	6
2	0	25	45	57	65	70	73
3	0	18	39	61	78	90	95
4	0	28	47	65	74	80	85

可以把对 4 个工厂的投资依次看成 4 个阶段的决策过程，确定对第 i 个工厂的投资额看成是第 i 个阶段的决策，$i=1,2,3,4$，如图 7.4 所示。第一个阶段开始时拥有的资金为 6 万元，即考虑第 1 个工厂投资决策时的资金数是 6 万元，投资第一个工厂的资金数为 u_1，所以在考虑第 2 个工厂投资时剩余的资金额为 $6-u_1$，以此类推，第 4 阶段的决策是将剩余资金都投入建厂，这样第 5 阶段的投资金额为 0，即 $s_5=0$。

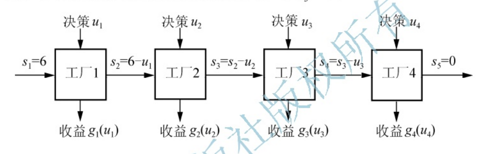


图 7.4　4 阶段动态规划问题

在多阶段决策过程中，动态规划是既把当前的一段和未来的各段分开，又把当前效益和未来效益结合起来考虑的一种最优化方法。动态规划主要用于求解以时间划分阶段的动态过程优化问题，但是一些与时间无关的静态规划(如线性规划、非线性规划)，只要人为地引进时间因素，将它视为多阶段决策过程，也可用动态规划方法方便地处理，如例 7.4。

7.1.2　动态规划问题的特点和分类

通常多阶段决策过程的发展是通过状态的一系列变换来实现的。一般情况下，系统在某个阶段的状态的决定除与本阶段的状态和决策有关外，还可能与系统过去经历的状态和决策有关，这样问题的求解就比较困难复杂。而适合于用动态规划方法求解的只是一类特殊的多阶段决策问题，即具有“无后效性”的多阶段决策过程。所谓无后效性，又称马尔科夫性，是指系统从某个阶段往后的发展，仅由本阶段所处的状态及其往后的决策所决定，与系统以前经历的状态和决策(历史)无关。

这个性质也可以形象地理解为动态规划问题结构上的可分性和自相似性。图 7.5(a)、(b)是图 7.3 的子部分，也是图 7.3 的子问题，在结构上和图 7.3 是相似的，解决这个子问题的思想、方法与图 7.3 所给的全局问题完全一致。后面讨论的多阶段决策问题都具有这种马尔科夫性，这里不再赘述。

在动态规划中，根据决策变量(决策方案)取值(连续或离散)的情况，将动态规划分为连续型、离散型动态规划；根据决策状态(确定或随机)，将动态规划分为确定型、随机型动态规划。这样动态规划有以下 4 种类型：离散确定型、离散随机型、连续确定型和连续随机型，其中应用最广的是确定型动态规划。

北京大学出版社版权所有 禁止转载

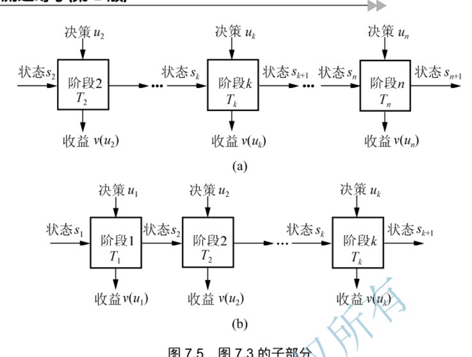


图 7.5　图 7.3 的子部分

北京大学出版社版权所有 禁止转载

7.2　动态规划的基本概念和数学模型

7.2.1　动态规划的基本概念

为描述动态规划的数学模型，先给出描述动态规划的有关概念。

1. 阶段与阶段变量

将所给动态规划问题的决策过程，划分为若干个相互联系的阶段，可以按照时间或空间的自然特征来分割，构建一个多阶段决策问题。描述决策阶段多少的变量叫做阶段变量，一般用 k 来表示，如例 7.1 的阶段可以按照时间安排为 4 个，阶段变量取值为 $k=1, 2, 3, 4$；例 7.4 的阶段可按照所投资的 4 个工厂的顺序划分为 4 个阶段，阶段变量取值也为 $k=1, 2, 3, 4$。

2. 状态与状态变量

状态表示每个阶段开始所处的自然状态和客观条件，也就是该阶段决策所依赖的初始条件。状态是决策的环境，是不可控因素。如例 7.3 各阶段的状态就是每一阶段的左端点，即各阶段的起始位置。每个阶段的状态可能不止一个，将第 k 阶段的状态集合用 S_k 表示，称为第 k 阶段可达状态集合。由于第三阶段有 3 个初始状态，则 $S_3=\{B_1, B_2, B_3\}$；第四阶段有 2 个初始状态，则 $S_4=\{C_1, C_2\}$。

描述状态的变量称为状态变量，第 k 阶段状态变量用小写 $s_k, k=1, 2, \cdots, n$ 表示。如例 7.3 处在状态 C_2 时，可表示为 $s_4=C_2$。

作为动态规划的状态变量，应满足无后效性(马尔科夫性)，之所以要求具有这种性质，是由于不具有无后效性的多阶段决策过程而言，不可能在不知道前面状态的情况下，逐段递推求解，因此对实际问题来说，必须正确地选择状态变量，使它所确定的过程具有无后效性，否则，就不能用来构造动态规划模型，并应用动态规划方法求解。

3. 决策与决策变量

当过程处于第k个阶段的某个状态时，可以作出不同的决定(选择)，从而确定下一阶段的状态，这种选择就是决策。描述决策方案的变量，称为决策变量，其可以是一个数，一组数，常用$u_k(s_k)$表示第k阶段处在状态s_k的决策变量，是状态变量的函数，在不会引起混淆的情况下，也可以简记为u_k。在实际问题中，决策变量的取值要受到已知条件的限制，取值在一定范围之内，将决策变量的取值范围称为允许决策集合，常用$D_k(s_k)$表示第k阶段处在状态s_k的决策方案集合，显然，$u_k(s_k)\in D_k(s_k)$。

例 7.2 处在状态C_2时，$D_3(C_2)=\{D_1,D_2\}$。

4. 策略与最优策略

动态规划的每一阶段都需要决策，则由第一阶段到第n阶段全过程的决策所构成的任一可行决策序列，称为一个策略，记为$p_{1n}(s_1)$或简记为p_{1n}，即

$$p_{1n}(s_1)=p_{1n}=\{u_1(s_1),u_2(s_2),\cdots,u_n(s_n)\}$$

从第k阶段到第n阶段的过程称为全过程的k后部子过程，其相应的决策序列称为k后部子策略，简记为p_{kn}，即

$$p_{kn}(s_k)=p_{kn}=\{u_k(s_k),u_{k+1}(s_{k+1}),\cdots,u_n(s_n)\}\quad k=1,2,\cdots,n$$

实际问题中，由于每个阶段都有若干个状态，针对每一个状态，又有不同的决策，从而组成了不同的决策函数序列，即存在许多策略可供选择。这种可供选择的策略范围，称为允许策略集合，记为P，从允许策略集合中找出使问题达到最优效果的策略称为最优策略，记作p_{1n}^*，即

$$p_{1n}^*(s_1)=p_{1n}^*=\{u_1^*(s_1),u_2^*(s_2),\cdots,u_n^*(s_n)\}$$

并称由第k阶段到第n阶段的最优策略为k后部最优子策略，记作$p_{kn}^*(s_k)$，即

$$p_{kn}^*(s_k)=p_{kn}^*=\{u_k^*(s_k),u_{k+1}^*(s_{k+1}),\cdots,u_n^*(s_n)\}$$

同样可以定义和描述动态规划的前部子策略和最优子策略，动态规划问题就是在允许策略集中选择最优策略。

5. 状态转移方程

多阶段决策过程是一个序贯决策过程，即如果已给定第k阶段的状态变量s_k，则在该阶段的决策变量u_k确定之后，则第$k+1$阶段的状态s_{k+1}也就随之确定，这样，可以把s_{k+1}看成是s_k、u_k的函数，即$s_{k+1}=T_k(s_k,u_k)$。这一关系指明了由第k阶段到第$k+1$阶段的状态转移规律，称为状态转移方程或状态转移函数。

例 7.4 的第 3 阶段到第 4 阶段的状态转移方程为$s_4=s_3-u_3$，其中u_3表示第 3 阶段所作的决策。

6. 指标函数与阶段指标函数

动态规划主要是用来解决多阶段过程最优化问题，因而要有一个用来衡量所实现过程优劣的一种数量指标，以便对某给定的策略进行评价，这就是指标函数。阶段指标函数(又称阶段效益)记为$v_k(s_k,u_k)$，它表示在第k阶段处于s_k状态下，经过决策u_k后在本阶段获得的收益。图 7.2 所示阶段指标就是相邻两点间的距离，$B_1\to C_2$的距离是 6。

北京大学出版社版权所有 禁止转载

过程指标函数(又称目标函数)是用来衡量过程所实现优劣的一种数量指标，它是定义在全过程和所有后部子过程上的数量函数，记作V_{kn}，即

$$V_{kn}(s_k)=V_{kn}(s_k,u_k,s_{k+1},u_{k+1},\cdots,s_n,u_n)\quad k=1,2,\cdots,n$$

其中u_k表示第k阶段的决策，s_{k+1}由s_k和u_k确定，其余同理，特别地，当$k=1$时就是全过程的指标函数。

当初始状态给定时，若过程的策略也确定了，则指标函数也就确定了，故指标函数也是初始状态和策略的函数，即$V_{kn}(s_k)=V_{kn}(s_k,p_{kn}(s_k))\quad k=1,2,\cdots,n$。

1) 由于动态规划方法是逐段递推求解的，作为动态规划模型的指标函数应满足的条件

(1) 指标函数应在全过程和所有后部子过程上有定义。

(2) 指标函数应具有可分离性。可分离性要求指标函数满足递推关系：对于任意k ($1\leqslant k\leqslant n$)，有

$$V_{kn}(s_k)=\psi_k(s_k,u_k,V_{k+1,n}(s_{k+1},u_{k+1},\cdots,s_n,u_n))$$

即指标函数V_{kn}可表达成第k段指标函数与从$k+1$段开始的后部子过程指标函数的一个函数关系，即V_{kn}能从函数关系ψ_k中由$V_{k+1,n}$和s_k,u_k得到。

(3) 函数ψ_k对其变元$V_{k+1,n}$来说要严格单调。条件(3)是为了保证目标函数在发展方向上的统一，不然目标就会发生混乱。这些要求是保证动态规划递推求解的关键，不过在实际应用中也不需考虑，因为以后用到的指标函数是最常见的两种——和与积指标函数，都满足上面的要求。

在例 7.3 中，如选择的策略是$Q\to A_1\to B_2\to C_1\to T$，则全过程的指标为各阶段的指标之和，即 3+4+2+3=12。选择的策略不同，全过程的指标函数值也不同，因而指标函数是全过程(或后部子过程)各阶段的状态和决策的函数。

2) 在实际问题中，常用的指标函数有两种基本形式

(1) 全过程和它的任一后部子过程的指标函数等于各阶段指标函数之和称为和指标函数，即

$$V_{kn}=\sum_{i=k}^{n}v_i(s_i,u_i)=v_k(s_k,u_k)+\sum_{i=k+1}^{n}v_i(s_i,u_i)=v_k(s_k,u_k)+V_{k+1,n}\quad k=1,2,\cdots,n$$

其中$v_k(s_k,u_k)$表示第i阶段的阶段指标函数。有时指标函数要表达为阶段初始状态和策略的函数，因而上式又可写成

$$V_{kn}(s_k,p_{kn}(s_k))=v_k(s_k,u_k)+V_{k+1,n}(s_{k+1},p_{k+1,n}(s_{k+1}))\quad k=1,2,\cdots,n$$

(2) 全过程和它的任一后部子过程的指标函数等于各阶段指标函数之积称为积指标函数，即

$$V_{kn}=\prod_{i=k}^{n}v_i(s_i,u_i)=v_k(s_k,u_k)\times\prod_{i=k+1}^{n}v_i(s_i,u_i)=v_k(s_k,u_k)\times V_{k+1,n}\quad k=1,2,\cdots,n$$

7. 最优值函数

指标函数V_{kn}的最优值，称为最优值函数，记为$f_k(s_k)$，它表示从第k阶段的状态s_k出发到过程结束时所获得的指标函数的最优值，即所有k后部子过程指标函数的最优值。

$$f_k(s_k)=\underset{u_k,\cdots,u_n}{\text{opt}}\{V_{kn}(s_k,u_k,\cdots,s_n,u_n)\}$$

北京大学出版社版权所有 禁止转载

其中 *opt* 是最优化(Optimization)的缩写，可根据题意而取 max 或 min。

从上面描述可知，最优值函数又可写成

$$f_k(s_k) = \underset{u_k,\cdots,u_n}{\text{opt}} \{V_{kn}(s_k, p_{kn}(s_k))\} = V_{kn}(s_k, p_{kn}^*(s_k))$$

其中，$p_{kn}^*(s_k)$ 表示初始状态为 s_k 时的后部子过程所有子策略中的最优子策略。

要特别留意最优值函数的符号表示，$f_k(s_k)$ 仅与其后部子过程的初始状态 s_k 有关，因为这个值(最优)可由初始状态完全确定，而指标函数除了依赖于初始状态，还取决于所选择的过程子策略。

7.2.2　动态规划的数学模型

有了前面的概念，就可以建立动态规划方法的数学模型为

$$f_1 = f_1(s_1) = \underset{u_1 \sim u_n}{\text{opt}} \{V_{1n}(s_1, u_1, s_2, u_2, \cdots, s_n, u_n)\}$$

$$\begin{cases} s_{k+1} = T_k(s_k, u_k) \\ s_k \in S_k \\ u_k \in U_k \\ k = 1, 2, \cdots, n \end{cases} \tag{7-1}$$

对于给定的多阶段决策过程，寻求一个(或多个)最优策略或最优决策序列 $u_1^*, u_2^*, \cdots, u_n^*$ 就是模型的解，使之既满足式(7-1)给出的全部约束条件，又使式(7-1)的目标函数取得极值，同时也能给出执行该策略时过程状态的演变序列。

北京大学出版社版权所有 禁止转载

7.3　动态规划的最优化原理和基本方程

根据上面的讨论，求动态规划模型的解，就是求指标函数的最优值，即由函数 $f_1(s_1)$ 确定。从 $f_1(s_1)$ 的表达式可知，是一个多变量的最优化问题，求解仍然困难复杂。如例 7.3，用完全枚举法，可供选择的路线有 40 条，将其一一比较才可找出最短路线，显然这样找法是不经济的，特别是当阶段数很多，各阶段可供的选择也很多时，这种解法甚至在计算机上完成也不现实。幸运的是，运筹大师 R. Bellman 等人根据多阶段决策问题的特点，提出了“最优化原理”作为动态规划的理论基础，它能解决许多类型的多阶段决策过程的最优化的问题。

7.3.1　Bellman 最优化原理

Bellman 最优化原理：“作为整个过程的最优策略具有这样的性质，即无论过去的状态和决策如何，对于先前的决策所形成的状态而言，余下的诸决策必须构成最优策略。”简言之，“一个最优策略的子策略总是最优的”。

基于 Bellman 最优化原理，考虑整个过程的最优策略，可以从子策略入手，从最优子策略中搜寻整个过程的最优策略，即从一个阶段的最优子策略，搜寻两个阶段的最优子策略，再搜寻三个阶段的最优策略，一直继续下去，就可以得到 n 个阶段的最优策略。

利用动态规划问题的特点，就可以按图 7.3 所示，将一个多阶段决策过程转化为一个连贯的决策过程，即把一个含有 n 个变量的决策问题转化为 n 个单变量决策问题。但要实

现这种转化，还要满足两个基本条件，这就是指标函数的可分性和状态变量的无后效性。

Bellman 最优化原理仅仅是策略最优性的必要条件，所以决定了动态规划求解思想是搜索过程，计算烦琐和复杂。不过计算机这个工具可以替代管理者做大量的重复工作，而他们只需熟悉解决问题的思路和途径。

7.3.2 动态规划的基本方程

在用动态规划方法去处理问题时，必须对实际问题建立起动态规划模型，明确变量和参数。与线性规划不同，动态规划模型没有一个统一的模式，它必须根据具体问题进行具体分析，在进行分析时，必须做到以下几点。

(1) 将实际问题恰当地划分为若干个阶段，一般是根据时间和空间的自然特性来划分，但要便于把问题的过程转化成多阶段决策的过程。

(2) 正确地选择状态变量 s_k，使它既能描述过程的演变特性，又要满足无后效性。

(3) 确定决策变量 u_k 及每阶段的允许决策集合 $D_k(s_k)$。

(4) 正确写出状态转移方程 $s_{k+1}=T_k(s_k,u_k)$。

(5) 正确写出阶段指标函数 $v_k(s_k,u_k)$ 和过程指标函数 V_{kn}。

上面 5 点是构造动态规划模型的基础，是正确写出动态规划基本方程的基本要素。一个问题的动态规划模型构造得是否合理，又集中地反映在要恰当地定义最优值函数、正确地写出递推关系和边界条件，下面就来讨论这个问题。

对于和指标函数而言，由最优值函数的定义可知

$$
\begin{aligned}
f_k(s_k) &= \operatorname*{opt}_{u_k,\cdots,u_n}\{V_{kn}(s_k,p_{kn}(s_k))\} \\
&= \operatorname*{opt}_{u_k,\cdots,u_n}\{v_k(s_k,u_k)+V_{k+1,n}(s_{k+1},p_{k+1,n}(s_{k+1}))\}
\end{aligned}
$$

其中，$s_{k+1}=T(s_k,u_k)$。因此

$$
\begin{aligned}
f_k(s_k) &= \operatorname*{opt}_{u_k,\cdots,u_n}\{v_k(s_k,u_k)+V_{k+1,n}(s_{k+1},p_{k+1,n}(s_{k+1}))\} \\
&= \operatorname*{opt}_{u_k,\,p_{k+1,n}}\{v_k(s_k,u_k)+V_{k+1,n}(s_{k+1},p_{k+1,n}(s_{k+1}))\} \\
&= \operatorname*{opt}_{u_k}\left\{v_k(s_k,u_k)+\operatorname*{opt}_{p_{k+1,n}}\{V_{k+1,n}(s_{k+1},p_{k+1,n}(s_{k+1}))\}\right\} \\
&= \operatorname*{opt}_{u_k}\{v_k(s_k,u_k)+f_{k+1}(s_{k+1})\}
\end{aligned}
$$

若能够得到最优值函数 $f_{k+1}(s_{k+1})$，那么通过上述公式就可以求出 $f_k(s_k)$，这样通过递推关系最终求出 $f_1=f_1(s_1)$，即整个问题的最优值，从而求出最优策略。为使递推过程顺利进行，还需添加边界条件 $f_{n+1}(s_{n+1})=0$。因为零元素在加法中不起作用，可以保证递推时不改变过程指标函数原来的数值，这样，就得到动态规划的基本方程或递推方程

$$
\begin{cases}
f_k(s_k)=\operatorname*{opt}_{u_k\in D_k(s_k)}\{v_k(s_k,u_k)+f_{k+1}(s_{k+1})\} \\
f_{n+1}(s_{n+1})=0 \\
s_{k+1}=T_k(s_k,u_k) \quad k=n,n-1,\cdots,1
\end{cases}
\tag{7-2}
$$

由于递推方程是从 $k=n$ 开始前往 $k=1$ 递推，方程(7-2)又称为逆序递推方程。在递推过程中，需将 s_{k+1} 用状态转移方程 $s_{k+1}=T_k(s_k,u_k)$ 去替换。

北京大学出版社版权所有
禁止转载

同理，对于积指标函数，也可写出它的一组基本方程

$$\begin{cases} f_k(s_k) = \underset{u_k \in D_k(s_k)}{\text{opt}} \{v_k(s_k, u_k) \times f_{k+1}(s_{k+1})\} \\ f_{n+1}(s_{n+1}) = 1 \\ s_{k+1} = T_k(s_k, u_k) \quad k = n, n-1, \cdots, 1 \end{cases} \tag{7-3}$$

递推求解的一般过程是：运用边界条件，从 $k = n$ 开始，由后向前逆推，从而逐步求得各阶段的最优决策和相应的最优值，最后求出 $f_1(s_1)$ 就是全过程的最优值，将 s_1 的值代入计算即得最优值，然后再由 s_1 和 u_1^* 利用状态转移方程计算出 s_2，从而确定 u_2^*，以此类推，最后确定 u_n^*，于是得最优策略 $p_{1n}^* = \{u_1^*, u_2^*, \cdots, u_n^*\}$。后面的计算过程称为“回代”，又称为“反向追踪”，动态规划的计算过程是由递推和回代两部分组成。

此外，动态规划解的寻找途径可以分为顺序和逆序两种方式。所谓顺序是指寻优过程与阶段进展的次序一致；所谓逆序是指寻优过程与阶段进展的次序相反。

与逆序递推方程推导的方法类似，设动态规划演化的终止状态为 s_{n+1}，并用 $f_k(s)$ 表示第 k 阶段末的结束状态为 s，从第一阶段到第 k 阶段所获得的最优值，其他符号的含义不变，可得顺序递推方程如下。

对于和指标函数，有

$$\begin{cases} f_k(s_{k+1}) = \underset{u_k \in D_k'(s_{k+1})}{\text{opt}} \{v_k(s_{k+1}, u_k) + f_{k-1}(s_k)\} \\ f_0(s_1) = 0 \\ s_k = T_k'(s_{k+1}, u_k) \quad k = 1, 2, \cdots, n \end{cases} \tag{7-4}$$

对于积指标函数，有

$$\begin{cases} f_k(s_{k+1}) = \underset{u_k \in D_k'(s_{k+1})}{\text{opt}} \{v_k(s_{k+1}, u_k) \times f_{k-1}(s_k)\} \\ f_0(s_1) = 1 \\ s_k = T_k'(s_{k+1}, u_k) \quad k = 1, 2, \cdots, n \end{cases} \tag{7-5}$$

其中：$s_k = T_k'(s_{k+1}, u_k)$ 是状态转移方程的变形，就是说要从 s_{k+1}, u_k 中反求 s_k，与前面给出的状态转移方程有所区别；$u_k \in D_k'(s_{k+1})$ 是指能到达状态 s_{k+1} 的所有第 k 阶段的决策；$v_k(s_{k+1}, u_k)$ 是经过决策 u_k 到达状态 s_{k+1} 的阶段指标函数，这些在顺序递推中需谨记。

北京大学出版社版权所有 禁止转载

从上面的描述不难得出：当阶段的初始状态给定时，用逆序的方式比较好；当终止状态给定时，用顺序的方式比较好。逆序解法表示较为符合思维习惯，表达较为简洁，实际中初始状态给定的情况居多，用逆序的方式也比较多。

从递推方程可以看出，动态规划就是这样一种在数学和计算机科学中使用的，用于求解包含重叠子问题的最优化问题的方法。动态规划在寻找有很多重叠子问题的情况的最优解时有效，为了避免多次解决这些子问题，它们的结果都逐渐被计算并被保存，从简单的问题直到整个问题都被解决。

动态规划的成功之处在于，把一个 n 阶段决策问题变换为 n 个单阶段最优化问题，一个一个地求解。这是经典极值方法所做不到的，它几乎超越了所有现存的计算方法，特别是经典优化方法。另外，动态规划能够求出全局极大或极小，这也是其他优化方法很难做到的。动态规划方法的基本思想体现了多阶段性、无后效性、递归性、全局优化性。动态规划是考察问题的一种途径，而不是一种特殊的算法，它不像线性规划那样有统一的数学模型和算法(例如单纯形法)，而必须对具体问题进行具体分析。针对不同的问题，运用动

态规划的原理和方法，建立起相应的模型，然后再用动态规划方法去求解。因此在学习时，除了要对动态规划的基本原理和方法正确理解外，还应以丰富的想象力去建立模型，用灵活的技巧去求解。

【例 7.5】 求解例 7.3 中的动态规划问题。

解：采用逆序和顺序求该动态规划问题的最优策略，逆序从 $T \to Q$ 递推，顺序从 $Q \to T$ 递推，显然这是一个四阶段动态规划问题。

1. *逆序法*

明确动态规划基本要素，建立动态规划基本方程。该题是一个最短路问题，其阶段指标表示相邻两个位置的距离，所以采用和指标函数，并求最小值，得

$$\begin{cases} f_k(s_k) = \min\limits_{u_k \in D_k(s_k)} \{v_k(s_k, u_k) + f_{k+1}(s_{k+1})\} \\ \quad f_5(T) = 0 \\ \quad s_{k+1} = T_k(s_k, u_k) \quad k = 4, 3, 2, 1 \end{cases}$$

1) 第四阶段($k = 4$)

此时 $S_5 = \{T\}$ 、 $S_4 = \{C_1, C_2\}$ ，从 C_1, C_2 到 T 都只有一条路线，将所选择路线的终点记为决策方案，因为这个记法可以追寻到所走的路线，即 $D_4(C_1) = \{T\}$ 、 $D_4(C_2) = \{T\}$ ，$v_4(C_1, T) = 3$ 、 $v_4(C_2, T) = 2$ ，则

$$f_4(C_1) = \min_{u_4 \in D_4(C_1)} \{v_4(C_1, u_4) + f_5(s_5)\} = \min_{u_4 \in D_4(C_1)} \{v_4(C_1, T) + f_5(T)\} = \min_{u_4 \in D_4(C_1)} \{3 + 0\} = 3$$

其相应的最优决策是 $u_4^*(C_1) = T$ ，这说明从 C_1 到 T 的最短路线是 $C_1 \to T$ ，最短路线的长度是 $f_4(C_1) = 3$ 。

同理，可以得

$$f_4(C_2) = \min_{u_4 \in D_4(C_2)} \{v_4(C_2, u_4) + f_5(s_5)\} = \min_{u_4 \in D_4(C_2)} \{v_4(C_1, T) + f_5(T)\} = \min_{u_4 \in D_4(C_2)} \{2 + 0\} = 2$$

其相应的最优决策是 $u_4^*(C_2) = T$ ，这说明从 C_2 到 T 的最短路线是 $C_2 \to T$ ，最短路线的长度是 $f_4(C_2) = 2$ 。

2) 第三阶段($k = 3$)

此时 $S_3 = \{B_1, B_2, B_3\}$, $D_3(B_1) = \{C_1, C_2\}$ 、$D_3(B_2) = \{C_1, C_2\}$ 、$D_3(B_3) = \{C_1, C_2\}$, $v_3(B_1, C_1) = 5$ 、$v_3(B_1, C_2) = 6$ 、 $v_3(B_2, C_1) = 2$ 、 $v_3(B_2, C_2) = 3$ 、 $v_3(B_3, C_1) = 2$ 、 $v_3(B_3, C_2) = 6$ ，则

$$f_3(B_1) = \min_{u_3 \in D_3(B_1)} \{v_3(B_1, u_3) + f_4(s_4)\} = \min_{u_3 \in D_3(B_1)} \begin{Bmatrix} v_3(B_1, C_1) + f_4(C_1) \\ v_3(B_1, C_2) + f_4(C_2) \end{Bmatrix} = \min_{u_3 \in D_3(B_1)} \begin{Bmatrix} 5 + 3 \\ 6 + 2 \end{Bmatrix} = 8$$

其相应的最优决策是 $u_3^*(B_1) = \{C_1, C_2\}$ ，这说明从 B_1 到 T 的最短路线有两条 $B_1 \to C_1 \to T$ 和 $B_1 \to C_2 \to T$ ，最短路线的长度是 $f_3(B_1) = 8$ 。

$$f_3(B_2) = \min_{u_3 \in D_3(B_2)} \{v_3(B_2, u_3) + f_4(s_4)\} = \min_{u_3 \in D_3(B_2)} \begin{Bmatrix} v_3(B_2, C_1) + f_4(C_1) \\ v_3(B_2, C_2) + f_4(C_2) \end{Bmatrix} = \min_{u_3 \in D_3(B_2)} \begin{Bmatrix} 2 + 3 \\ 3 + 2 \end{Bmatrix} = 5$$

其相应的最优决策是 $u_3^*(B_2) = \{C_1, C_2\}$ ，这说明从 B_2 到 T 的最短路线是 $B_2 \to C_1 \to T$ 和 $B_2 \to C_2 \to T$ ，最短路线的长度是 $f_3(B_2) = 5$ 。

$$f_3(B_3) = \min_{u_3 \in D_3(B_3)} \{v_3(B_3, u_3) + f_4(s_4)\} = \min_{u_3 \in D_3(B_3)} \begin{Bmatrix} v_3(B_3, C_1) + f_4(C_1) \\ v_3(B_3, C_2) + f_4(C_2) \end{Bmatrix} = \min_{u_3 \in D_3(B_3)} \begin{Bmatrix} 2 + 3 \\ 6 + 2 \end{Bmatrix} = 5$$

北京大学出版社版权所有
禁止转载

其相应的最优决策是 $u_3^*(B_3)=\{C_1\}$，这说明从 B_3 到 T 的最短路线是 $B_3\to C_1\to T$，最短路线的长度是 $f_3(B_3)=5$。

3) 第二阶段($k=2$)

此时 $S_2=\{A_1,A_2,A_3\}$，$D_2(A_1)=\{B_1,B_2,B_3\}$、$D_2(A_2)=\{B_1,B_3\}$、$D_2(A_3)=\{B_2,B_3\}$，$v_2(A_1,B_1)=7$、$v_2(A_1,B_2)=4$、$v_2(A_1,B_3)=6$、$v_2(A_2,B_1)=4$、$v_2(A_2,B_3)=1$、$v_2(A_3,B_2)=3$、$v_2(A_3,B_3)=2$，则

$$f_2(A_1)=\min_{u_2\in D_2(A_1)}\{v_2(A_1,u_2)+f_3(s_3)\}=\min_{u_2\in D_2(A_1)}\begin{Bmatrix}v_2(A_1,B_1)+f_3(B_1)\\v_2(A_1,B_2)+f_3(B_2)\\v_2(A_1,B_3)+f_3(B_3)\end{Bmatrix}=\min_{u_2\in D_2(A_1)}\begin{Bmatrix}7+8\\4+5\\6+5\end{Bmatrix}=9$$

其相应的最优决策是 $u_2^*(A_1)=\{B_2\}$，这说明从 A_1 到 T 的最短路线经过 B_2，最短路线的长度是 $f_2(A_1)=9$。

$$f_2(A_2)=\min_{u_2\in D_2(A_2)}\{v_2(A_2,u_2)+f_3(s_3)\}=\min_{u_2\in D_2(A_2)}\begin{Bmatrix}v_2(A_2,B_1)+f_3(B_1)\\v_2(A_2,B_3)+f_3(B_3)\end{Bmatrix}=\min_{u_2\in D_2(A_2)}\begin{Bmatrix}4+8\\1+5\end{Bmatrix}=6$$

其相应的最优决策是 $u_2^*(A_2)=\{B_3\}$，这说明从 A_2 到 T 的最短路线经过 B_3，最短路线的长度是 $f_2(A_2)=6$。

$$f_2(A_3)=\min_{u_2\in D_2(A_3)}\{v_2(A_3,u_2)+f_3(s_3)\}=\min_{u_2\in D_2(A_3)}\begin{Bmatrix}v_2(A_3,B_2)+f_3(B_2)\\v_2(A_3,B_3)+f_3(B_3)\end{Bmatrix}=\min_{u_2\in D_2(A_3)}\begin{Bmatrix}3+5\\2+5\end{Bmatrix}=7$$

其相应的最优决策是 $u_2^*(A_3)=\{B_3\}$，这说明从 A_3 到 T 的最短路线经过 B_3，最短路线的长度是 $f_2(A_3)=7$。

4) 第一阶段($k=1$)

此时 $S_1=\{Q\}$，$D_1(Q)=\{A_1,A_2,A_3\}$，$v_1(Q,A_1)=3$、$v_1(Q,A_2)=5$、$v_1(Q,A_3)=4$，则

$$f_1(Q)=\min_{u_1\in D_1(Q)}\{v_1(Q,u_1)+f_2(s_2)\}=\min_{u_1\in D_1(Q)}\begin{Bmatrix}v_1(Q,A_1)+f_2(A_1)\\v_1(Q,A_2)+f_2(A_2)\\v_1(Q,A_3)+f_2(A_3)\end{Bmatrix}=\min_{u_1\in D_1(Q)}\begin{Bmatrix}3+9\\5+6\\4+7\end{Bmatrix}=11$$

其相应的最优决策是 $u_1^*(Q)=\{A_2,A_3\}$，这说明从 Q 到 T 的最短路线经过 A_2 或 A_3，最短路线的长度是 $f_1(Q)=11$。

最后逐段追溯回去，可知最短路径有两条：$Q\to A_2\to B_3\to C_1\to T$ 和 $Q\to A_3\to B_3\to C_1\to T$，得到两个最优策略，其最短路线的长度为 11。

2. 顺序法

根据式(7-4)，建立动态规划基本方程，得

$$\begin{cases}f_k(s_{k+1})=\min\limits_{u_k\in D_k'(s_{k+1})}\{v_k(s_{k+1},u_k)+f_{k-1}(s_k)\}\\ \quad f_0(Q)=0\\ \quad s_k=T_k'(s_{k+1},u_k)\quad k=1,2,3,4\end{cases}$$

有关数据在逆序法已经用到，下面简要给出计算过程。

1) 第一阶段($k=1$)

此时 $S_1=\{Q\}$、$S_2=\{A_1,A_2,A_3\}$，为便于追寻所走的路线，将所有路线的起点记为决策

北京大学出版社版权所有 禁止转载

方案，因为在第一阶段到达 A_1 只有 $Q \to A_1$ 一条路线，即 $D_1'(A_1)=\{Q\}$，则

$$f_1(A_1)=\min_{u_1 \in D_1'(A_1)}\{v_1(A_1,u_1)+f_0(s_1)\}=\min_{u_1 \in D_1'(A_1)}\{v_1(A_1,Q)+f_0(Q)\}=\min_{u_1 \in D_1'(A_1)}\{3+0\}=3$$

其相应的最优决策是 $u_1^*(A_1)=\{Q\}$ (到达状态 A_1 所采取的决策)，这说明从 Q 到 A_1 的最短路线经过 Q，最短路线的长度是 $f_1(A_1)=3$。

$$f_1(A_2)=\min_{u_1 \in D_1'(A_2)}\{v_1(A_2,u_1)+f_0(s_1)\}=\min_{u_1 \in D_1'(A_2)}\{v_1(A_2,Q)+f_0(Q)\}=\min_{u_1 \in D_1'(A_2)}\{5+0\}=5$$

其相应的最优决策是 $u_1^*(A_2)=\{Q\}$ (到达状态 A_2 所采取的决策)，这说明从 Q 到 A_2 的最短路线经过 Q，最短路线的长度是 $f_1(A_2)=5$。

$$f_1(A_3)=\min_{u_1 \in D_1'(A_3)}\{v_1(A_3,u_1)+f_0(s_1)\}=\min_{u_1 \in D_1'(A_3)}\{v_1(A_3,Q)+f_0(Q)\}=\min_{u_1 \in D_1'(A_3)}\{4+0\}=4$$

其相应的最优决策是 $u_1^*(A_3)=\{Q\}$ (到达状态 A_3 所采取的决策)，这说明从 Q 到 A_3 的最短路线经过 Q，最短路线的长度是 $f_1(A_3)=4$。

2) 第二阶段($k=2$)

此时 $S_3=\{B_1,B_2,B_3\}$，根据递推方程，得

$$f_2(B_1)=\min_{u_2 \in D_2'(B_1)}\{v_2(B_1,u_2)+f_1(s_2)\}=\min_{u_2 \in D_2'(B_1)}\begin{Bmatrix} v_2(B_1,A_1)+f_1(A_1) \\ v_2(B_1,A_2)+f_1(A_2) \end{Bmatrix}=\min_{u_2 \in D_2'(B_1)}\begin{Bmatrix} 7+3 \\ 4+5 \end{Bmatrix}=9$$

其相应的最优决策是 $u_2^*(B_1)=\{A_2\}$，这说明从 Q 到 B_1 的最短路线经过 A_2，最短路线的长度是 $f_2(B_1)=9$。

同理，可知

$$f_2(B_2)=\min_{u_2 \in D_2'(B_2)}\{v_2(B_2,u_2)+f_1(s_2)\}=\min_{u_2 \in D_2'(B_2)}\begin{Bmatrix} v_2(B_2,A_1)+f_1(A_1) \\ v_2(B_2,A_3)+f_1(A_3) \end{Bmatrix}=\min_{u_2 \in D_2'(B_2)}\begin{Bmatrix} 4+3 \\ 3+4 \end{Bmatrix}=7$$

其相应的最优决策是 $u_2^*(B_2)=\{A_1,A_2\}$，这说明从 Q 到 B_2 的最短路线经过 A_1 或 A_2，最短路线的长度是 $f_2(B_2)=7$。

$$f_2(B_3)=\min_{u_2 \in D_2'(B_3)}\{v_2(B_3,u_2)+f_1(s_2)\}=\min_{u_2 \in D_2'(B_3)}\begin{Bmatrix} v_2(B_3,A_1)+f_1(A_1) \\ v_2(B_3,A_2)+f_1(A_2) \\ v_2(B_3,A_3)+f_1(A_3) \end{Bmatrix}=\min_{u_2 \in D_2'(B_3)}\begin{Bmatrix} 6+3 \\ 1+5 \\ 2+4 \end{Bmatrix}=6$$

其相应的最优决策是 $u_2^*(B_3)=\{A_2,A_3\}$，这说明从 Q 到 B_3 的最短路线经过 A_2 或 A_3，最短路线的长度是 $f_2(B_3)=6$。

3) 第三阶段($k=3$)

此时 $S_4=\{C_1,C_2\}$，根据递推方程，得

$$f_3(C_1)=\min_{u_3 \in D_3'(C_1)}\{v_3(C_1,u_3)+f_2(s_3)\}=\min_{u_3 \in D_3'(C_1)}\begin{Bmatrix} v_3(C_1,B_1)+f_2(B_1) \\ v_3(C_1,B_2)+f_2(B_2) \\ v_3(C_1,B_3)+f_2(B_3) \end{Bmatrix}=\min_{u_3 \in D_3'(C_1)}\begin{Bmatrix} 5+9 \\ 2+7 \\ 2+6 \end{Bmatrix}=8$$

其相应的最优决策是 $u_3^*(C_1)=\{B_3\}$，这说明从 Q 到 C_1 的最短路线经过 B_3，最短路线的长度是 $f_3(C_1)=8$。

$$f_3(C_2)=\min_{u_3 \in D_3'(C_2)}\{v_3(C_2,u_3)+f_2(s_3)\}=\min_{u_3 \in D_3'(C_2)}\begin{Bmatrix} v_3(C_2,B_1)+f_2(B_1) \\ v_3(C_2,B_2)+f_2(B_2) \\ v_3(C_2,B_3)+f_2(B_3) \end{Bmatrix}=\min_{u_3 \in D_3'(C_2)}\begin{Bmatrix} 6+9 \\ 3+7 \\ 6+6 \end{Bmatrix}=10$$

北京大学出版社版权所有
禁止转载

其相应的最优决策是$u_3^*(C_2)=\{B_2\}$，这说明从Q到C_2的最短路线经过B_2，最短路线的长度是$f_3(C_2)=10$。

4) 第四阶段($k=4$)

此时$S_5=\{T\}$，根据递推方程，得

$$f_4(T)=\min_{u_4\in D_4'(T)}\{v_4(T,u_4)+f_3(s_4)\}=\min_{u_4\in D_4'(T)}\begin{Bmatrix}v_4(T,C_1)+f_3(C_1)\\v_4(T,C_2)+f_3(C_2)\end{Bmatrix}=\min_{u_4\in D_4'(T)}\begin{Bmatrix}3+8\\2+10\end{Bmatrix}=11$$

其相应的最优决策是$u_4^*(T)=\{C_1\}$，这说明从Q到T的最短路线经过C_1，最短路线的长度是$f_4(T)=11$。

最后逐段追溯回去，可知最短路径有两条：$T\to C_1\to B_3\to A_2\to Q$和$T\to C_1\to B_3\to A_3\to Q$，有两个最优策略，其最短路线的长度为 11，如图 7.6 的箭线所指线路。

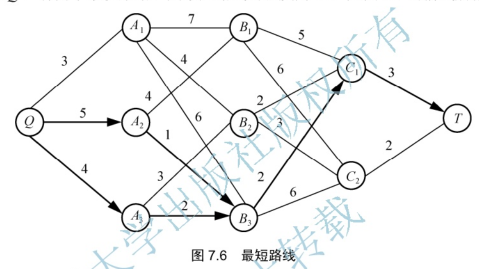


图 7.6　最短路线

北京大学出版社版权所有 禁止转载

从上述内容可知，顺序法与逆序解法的结果完全一样，但递推路线相反，符号不太习惯，建议以后尽量用逆序递推的方法求解动态规划。

【例 7.6】 求解例 7.2 中的动态规划问题，如图 7.7 所示。

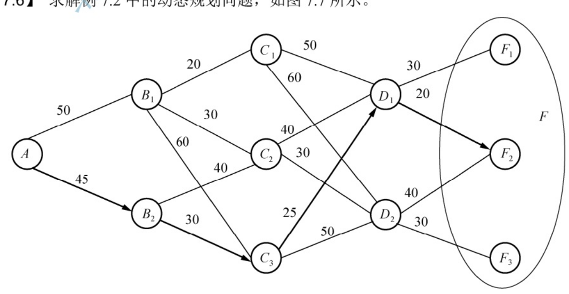


图 7.7　例 7.6 图

解：应用逆序解法求解从终点F向起点A的最优路线。

考虑到该题的特殊性，第四阶段结束状态不唯一(3 个港口)，造成边界条件设置上的困难，因此在建立动态规划模型时首先要处理这个难点。处理的方法很多，一是将该问题拆分为 3 个动态规划问题，F_1、F_2 和 F_3 各作为一个目的码头讨论其最优策略，最后比较这 3 个策略的最短路，获得原问题的最优策略；二是引入一个新的虚拟码头，使得其与 F_1、F_2 和 F_3 距离为零，增加一个阶段求解该问题，这样也不改变问题的最优路线；三是将 F_1、F_2 和 F_3 都作为终点，设其初始条件为零，求解该问题。为了简便求解，这里选用第三种方法，设置 3 个初始条件。

明确该问题的基本要素，这是一个四阶段动态规划问题，其阶段指标表示相邻两个位置的距离，采用和指标函数考量最优策略，为求最短路径，建立动态规划基本方程，得

$$\begin{cases} f_k(s_k) = \min\limits_{u_k \in D_k'(s_k)} \{v_k(s_k, u_k) + f_{k+1}(s_{k+1})\} \\ f_0(F_i) = 0 \qquad i = 1, 2, 3 \\ s_{k+1} = T_k(s_k, u_k) \quad k = 4, 3, 2, 1 \end{cases}$$

由于该问题的状态和决策的取值都是离散的，为便于叙述，特引入表格将相关要素组织在一起，计算各阶段的最优值，在解决实际问题中可模仿应用。

1) 第四阶段($k=4$)(见表 7-3)

表 7-3 第四阶段

s_4	$D_4(s_4)$	s_5	$v_4(s_4, u_4)$	$v_4(s_4, u_4) + f_5(s_5)$	$f_4(s_4)$	最优决策 $u_4(s_4)$
D_1	$D_1 \to F_1$ $D_1 \to F_2$	F_1 F_2	30 20	30+0 20+0	20	$D_1 \to F_2$
D_2	$D_2 \to F_2$ $D_2 \to F_3$	F_2 F_3	40 30	40+0 30+0	30	$D_2 \to F_3$

2) 第三阶段($k=3$)(见表 7-4)

表 7-4 第三阶段

s_3	$D_3(s_3)$	s_4	$v_3(s_3, u_3)$	$v_3(s_3, u_3) + f_4(s_4)$	$f_3(s_3)$	最优决策 $u_3(s_3)$
C_1	$C_1 \to D_1$ $C_1 \to D_2$	D_1 D_2	50 60	50+20 60+30	70	$C_1 \to D_1$
C_2	$C_2 \to D_1$ $C_2 \to D_2$	D_1 D_2	40 30	40+20 30+30	60	$C_2 \to D_1$ $C_2 \to D_2$
C_3	$C_3 \to D_1$ $C_3 \to D_2$	D_1 D_2	25 50	25+20 50+30	45	$C_3 \to D_1$

3) 第二阶段($k=2$)(见表 7-5)

表 7-5 第二阶段

s_2	$D_2(s_2)$	s_3	$v_2(s_2, u_2)$	$v_2(s_2, u_2) + f_3(s_3)$	$f_2(s_2)$	最优决策 $u_2(s_2)$
B_1	$B_1 \to C_1$ $B_1 \to C_2$ $B_1 \to C_3$	C_1 C_2 C_3	20 30 60	20+70 30+60 60+45	90	$B_1 \to C_1$ $B_1 \to C_2$

北京大学出版社版权所有 禁止转载

续表

s_2	$D_2(s_2)$	s_3	$v_2(s_2,u_2)$	$v_2(s_2,u_2)+f_3(s_3)$	$f_2(s_2)$	最优决策 $u_2(s_2)$
B_2	$B_2 \to C_2$	C_2	40	40+60	75	$B_2 \to C_3$
	$B_2 \to C_3$	C_3	30	30+45		

4) 第一阶段($k=1$)(见表 7-6)。

表 7-6　第一阶段

s_1	$D_1(s_1)$	s_2	$v_1(s_1,u_1)$	$v_1(s_1,u_1)+f_2(s_2)$	$f_1(s_1)$	最优决策 $u_1(s_1)$
A	$A \to B_1$	B_1	50	50+90	120	$A \to B_2$
	$A \to B_2$	B_2	45	45+75		

从上面结果得出，A 到 F 距离最短的是 120，最优路线为 $A \to B_2 \to C_3 \to D_1 \to F_2$，如图 7.7 中箭头示意路径。

北京大学出版社版权所有 禁止转载

7.4　动态规划的应用

求解一个标准的动态规划框架，通常要求按以下几个步骤进行：划分阶段、选择状态、确定决策并写出状态转移方程、写出递推方程(包括边界条件)。动态规划应用的主要难点在于框架上的设计，一旦设计完成，实现部分就会非常简单。根据动态规划的基本方程可以直接递推计算其最优值。由于动态规划模型与线性规划模型不同，动态规划模型没有统一的模式，建模时必须根据具体问题具体分析，只有通过不断实践总结，才能较好掌握建模方法与技巧。

7.4.1　生产库存问题

【例 7.7】某工厂制订一种产品的年度生产(或采购)计划，预测不同季度产品的市场需求分别为 2、3、2、4 单位，单位产品的固定成本 3 千元，可变成本 1 千元，每季度最大产品生产能力 $m=6$ 单位，每单位产品的季度存储费用 0.5 千元，库存容量不限，要求每年末库存为 0。试问：如何安排生产与库存，才能使得总成本最小？

解：明确动态规划的基本要素，建立动态规划递推方程。

按季度将该问题分为 4 阶段，$k=1,2,3,4$；

状态变量 s_k 为第 k 季度的库存数；

决策变量 x_k 为第 k 季度的产量，需求量为 d_k；

状态转移方程：$s_k = s_{k-1} + x_k - d_k$

阶段指标函数为季度的生产成本和存储费用：$v_k(s_k,x_k)=c_k(x_k)+h_k(s_k)$

其中生产成本：$c_k(x_k)=\begin{cases}0 & \text{当}x_k=0\\ 3+x_k & \text{当}x_k=1,\cdots,6\\ \infty & \text{当}x_k>6\end{cases}$

存储费用：$h_k(s_k)=0.5s_k$

动态规划的递推方程为

$$\begin{cases} f_k(s_k) = \min\limits_{u_k \in D_k(s_k)} \{v_k(s_k, x_k) + f_{k+1}(s_{k+1})\} \\ f_5(s_5) = 0 \quad k = 1, 2, 3, 4 \end{cases}$$

从边界条件出发，对于每一阶段 k，计算各状态(库存水平)的总成本 $f_k(s_k)$，其中，第 k 阶段的库存上限为 $s_k \leqslant \max\left\{\sum_{j=k}^{n} d_j, m - d_k\right\}$，这是因为第 k 阶段能够用于库存的数量为最大产量减去本阶段的需求量，即 $m - d_k$，同时其库存数量也不超过后续所有阶段的总需求 $\sum_{j=k}^{n} d_j$。

1) 第四阶段($k = 4$)

到第四阶段时，初始库存为 s_4，根据题知 $s_4 \leqslant 4$，$s_5 = 0$。如果 s_4 大于 4，就是第四季度不生产，到第四季度末库存仍不为零，不符题意，迭代过程见表 7-7(a)。

表 7-7(a) 第四阶段

s_4	$D_4(s_4)$	s_5	$v_4(s_4,x_4)$	$v_4(s_4,x_4)+f_5(s_5)$	$f_4(s_4)$	最优决策 x_4
0	4	0	7+0.5×0	7+0	7	4
1	3	0	6+0.5×1	6.5+0	6.5	3
2	2	0	5+0.5×2	6+0	6	2
3	1	0	4+0.5×3	5.5+0	5.5	1
4	0	0	0+0.5×4	2+0	2	0

2) 第三阶段($k = 3$)

到第三阶段时，初始库存为 s_3，因为每季度生产最大量都大于需求，所以其第三季度开始时的最大库存应满足第三季度和第四季度的需求，即 $s_3 \leqslant 6$，迭代过程见表 7-7(b)。

表 7-7(b) 第三阶段

s_3	$D_3(s_3)$	s_4	$v_3(s_3,x_3)$	$v_3(s_3,x_3)+f_4(s_4)$	$f_3(s_3)$	最优决策 x_3
0	2	0	5+0×0.5	5+7	11	6
	3	1	6+0×0.5	6+6.5		
	4	2	7+0×0.5	7+6		
	5	3	8+0×0.5	8+5.5		
	6	4	9+0×0.5	9+2		
1	1	0	4+1×0.5	4.5+7	10.5	5
	2	1	5+1×0.5	5.5+6.5		
	3	2	6+1×0.5	6.5+6		
	4	3	7+1×0.5	7.5+5.5		
	5	4	8+1×0.5	8.5+2		
2	0	0	0+2×0.5	1+7	8	0
	1	1	4+2×0.5	5+6.5		
	2	2	5+2×0.5	6+6		
	3	3	6+2×0.5	7+5.5		
	4	4	7+2×0.5	8+2		

北京大学出版社版权所有
禁止转载

续表

s_3	$D_3(s_3)$	s_4	$v_3(s_3,x_3)$	$v_3(s_3,x_3)+f_4(s_4)$	$f_3(s_3)$	最优决策 x_3
3	0	1	0+3×0.5	1.5+6.5	8	0
	1	2	4+3×0.5	5.5+6		
	2	3	5+3×0.5	6.5+5.5		
	3	4	6+3×0.5	7.5+2		
4	0	2	0+4×0.5	2+6	8	0
	1	3	4+4×0.5	6+5.5		
	2	4	5+4×0.5	7+2		
5	0	3	0+5×0.5	2.5+5.5	8	0
	1	4	4+5×0.5	6.5+2		
6	0	4	0+6×0.5	3+2	5	0

3) 第二阶段($k=2$)

第二阶段开始时的库存为s_2，考虑到第一个季度最多生产 6 个单位，满足第一个季度的需求后，得$s_2 \leqslant 4$，迭代过程见表 7-7(c)。

北京大学出版社版权所有
禁止转载

表 7-7(c)　第二阶段

s_2	$D_2(s_2)$	s_3	$v_2(s_2,x_2)$	$v_2(s_2,x_2)+f_3(s_3)$	$f_2(s_2)$	最优决策 x_2
0	3	0	6+0×0.5	6+11	16	5
	4	1	7+0×0.5	7+10.5		
	5	2	8+0×0.5	8+8		
	6	3	9+0×0.5	9+8		
1	2	0	5+1×0.5	5.5+11	15.5	4
	3	1	6+1×0.5	6.5+10.5		
	4	2	7+1×0.5	7.5+8		
	5	3	8+1×0.5	8.5+8		
	6	4	9+1×0.5	9.5+8		
2	1	0	4+2×0.5	5+11	15	3
	2	1	5+2×0.5	6+10.5		
	3	2	6+2×0.5	7+8		
	4	3	7+2×0.5	8+8		
	5	4	8+2×0.5	9+8		
	6	5	9+2×0.5	10+8		
3	0	0	0+3×0.5	1.5+11	12.5	0
	1	1	4+3×0.5	5.5+10.5		
	2	2	5+3×0.5	6.5+8		
	3	3	6+3×0.5	7.5+8		
	4	4	7+3×0.5	8.5+8		
	5	5	8+3×0.5	9.5+8		
	6	6	9+3×0.5	10.5+5		
4	0	1	0+4×0.5	2+10.5	12.5	0
	1	2	4+4×0.5	6+8		
	2	3	5+4×0.5	7+8		
	3	4	6+4×0.5	8+8		
	4	5	7+4×0.5	9+8		
	5	6	8+4×0.5	10+5		

4) 第一阶段($k=1$)

第一阶段开始时的库存为$s_1=0$，迭代过程见表 7-7(d)。

表 7-7(d) 第一阶段

s_1	$D_1(s_1)$	s_2	$v_1(s_1,x_1)$	$v_1(s_1,x_1)+f_2(s_2)$	$f_1(s_1)$	最优决策 x_1
0	2	0	5+0×0.5	5+16	20.5	5
	3	1	6+0×0.5	6+15.5		
	4	2	7+0×0.5	7+15		
	5	3	8+0×0.5	8+12.5		
	6	4	9+0×0.5	9+12.5		

逐级向后追溯，可得各阶段的最优决策：$x_1=5, x_2=0, x_3=6, x_4=0$。

7.4.2 资源分配问题

资源分配问题就是将一定数量的一种或若干种资源(原材料、资金、设备等)合理分配给若干使用者，使得资源分配后总效果最优。

【例 7.8】求解例 7.4 中的动态规划问题。

解：把对 4 个工厂的投资依次看成 4 个阶段的决策过程，将第k个工厂的投资额看成第k个阶段的决策，$k=1,2,3,4$。

状态变量s_k：第k阶段决策时剩余的资金额。

决策变量x_k：第k阶段对第k个工厂的投资额，允许决策集$D_k=\{0,1,\cdots,s_k\}$。

状态转移方程：$s_{k+1}=s_k-x_k$，$k=1,2,3,4$，其中$s_1=6$万元。

阶段指标函数$g_k(x_k)$：第k阶段投资x_k万元时的投资效益，见表 7-2。

最优指标函数$f_k(s_k)$：第k阶段状态为s_k且对第k个工厂投资为x_k时，第k个工厂以及以后产生的总投资效益。

基本递推方程为

$$\begin{cases} f_k(s_k)=\max\limits_{0\leqslant x_k\leqslant s_k}\{g_k(x_k)+f_k(s_{k+1})\} & k=4,3,2,1 \\ f_5(s_5)=0 \end{cases}$$

用逆序法求解，可得 4 个最优策略：

$$x_1=0,\quad x_2=2,\quad x_3=3,\quad x_4=1;$$
$$x_1=2,\quad x_2=1,\quad x_3=2,\quad x_4=1;$$
$$x_1=1,\quad x_2=1,\quad x_3=3,\quad x_4=1;$$
$$x_1=2,\quad x_2=2,\quad x_3=0,\quad x_4=2。$$

最大投资效益为 $f_1(S_1)$=$f_1(6)$=134 万吨。

【例 7.9】 某公司有 1 000 辆功能完好的运输卡车，在超负荷运输(即每天满载行驶 500km 以上)情况下，年利润为 25 万元/辆，这时卡车的年损坏率为 0.3；在低负荷运输(即每天行驶 300km 以下)情况下，年利润为 16 万元/辆，年损坏率为 0.1。现要制订一个 5 年计划，问每年年初应如何分配完好车辆(在两种不同的负荷下运输)，使在 5 年内的总利润最大？

解：这是一个以时间为特征的多阶段决策问题。

北京大学出版社版权所有 禁止转载

阶段：将 5 年运输计划看成 5 个阶段的决策问题，$k=1,2,3,4,5$。

状态变量 s_k：第 k 阶段初完好的卡车数量，$s_1=1\,000$。

决策变量 x_k：表示第 k 阶段用于分配给超负荷运输的卡车数量，显然，分配给低负荷的卡车数为 s_k-x_k。

状态转移方程：$s_{k+1}=(1-0.3)x_k+(1-0.1)(s_k-x_k)=0.9s_k-0.2x_k$。

阶段指标函数 $g_k(x_k)$：表示第 k 年度利润，$g_k(x_k)=25x_k+16(s_k-x_k)=16s_k+9x_k$。

最优指标函数 $f_k(s_k)$：第 k 年度初完好车辆数为 s_k 时，采用最优策略到第 5 年末所产生的最大利润，递推方程为

$$\begin{cases} f_k(s_k)=\max\limits_{0\leqslant x_k\leqslant s_k}\{g_k(x_k)+f_{k+1}(s_{k+1})\} & k=5,4,3,2,1 \\ f_6(s_6)=0 \end{cases}$$

用逆序法求解，可得最优策略：$x_1=0,\quad x_2=0,\quad x_3=810,\quad x_4=567,\quad x_5=396.9$，即第一年年初，1 000 辆车全部用于低负荷运输；第二年年初，还有 900 辆完好的车，也全部用于低负荷运输；第三年年初，还有 810 辆完好的车，全部用于超负荷运输；第四年年初：还有 567 辆完好的车，全部用于超负荷运输；第五年年初，还有 396.9 辆完好的车，全部用于超负荷运输；到第五年年末，即第六年年初，还剩余 277.83 辆完好的车。所创最大利润　　$f_1(s_1)=f_1(1\,000)$=74 747.5 万元。

【例 7.10】 一般的资源分配问题：资源数量为 a，将其分配给 n 个使用者，分配给第 i 个使用者数量 x_i 时，相应的收益为 $g_i(x_i)$，问如何分配使得总收益最大？该问题的数学模型为

$$\max z=g_1(x_1)+g_2(x_2)+\cdots+g_n(x_n)$$

$$\begin{cases} x_1+x_2+\cdots+x_n=a \\ x_i\geqslant 0 \quad i=1,2,\cdots,n \end{cases}$$

解：该问题是一个线性规划问题，应用动态规划方法求解时可人为地赋予时间概念，将其看做是一个多阶段决策问题。

按变量个数划分阶段，$k=1,2,\cdots,n$。

状态变量为 s_k：表示分配给第 k 个至第 n 个使用者的总资源数量。

决策变量 x_k：表示分配给第 k 个使用者的资源数量，允许决策集合

$$D_k(s_k)=\{x_k\mid 0\leqslant x_k\leqslant s_k\}$$

状态转移方程：$s_{k+1}=s_k-x_k$，其中 $s_1=a$。

阶段指标函数 $g_k(x_k)$：表示分配给第 k 个使用者 x_k 数量时的收益。

最优指标函数 $f_k(s_k)$：表示以 s_k 数量的资源分配给第 k 个至第 n 个使用者所得到的最大收益，则动态规划基本方程为

$$\begin{cases} f_k(s_k)=\max\limits_{0\leqslant x_k\leqslant s_k}[g_k(x_k)+f_{k+1}(s_{k+1})] & k=n,\cdots,1 \\ f_{n+1}(s_{n+1})=0 \end{cases}$$

由后向前逐段递推，$f_1(a)$ 即为所求问题的最大收益。

7.4.3　系统可靠性问题

【例 7.11】 考虑设计一种有 3 个主要元件的电气设备。这 3 个元件是串联的，所以一个元件

北京大学出版社版权所有 禁止转载

发生故障将使整个设备工作不能正常。设备的可靠性可以通过在每个元件上并联(备用)元件改进，每个元件可以由 1 个、2 个或 3 个元件并联。设计这种设备的总资金为 10 万元，对元件 $i(i=1,2,3)$ 配备 $m_i(m_i=1,2,3)$ 个元件后的可靠性为 $R_i(m_i)$，成本为 $C_i(m_i)$，具体数值由表 7-8(a)给出。试确定各元件并联的元件数目，使得在不超过总设计费用的前提下，设备的可靠性最大。

表 7-8(a)　设备成本与可靠性表　　单位：万元

m_i	i=1		i=2		i=3	
	R	C	R	C	R	C
1	0.6	1	0.7	3	0.5	2
2	0.8	2	0.8	5	0.7	4
3	0.9	3	0.9	6	0.9	5

解：设备的可靠性 $R=R_1(m_1)R_2(m_2)R_3(m_3)$，其中 $R_i(m_i)$ 是第 i 个元件有 m_i 个并联元件时的可靠性。

阶段变量 $k=1,2,3$ 表示元件编号；状态变量 s_k 表示设计第 k 个元件前剩余的资金数，即第 k 个元件到第 n 个元件的总费用；决策变量 u_k 表示第 k 个元件的并联元件数。状态转移方程为 $s_{k+1}=s_k-C_k(u_k)$，基本递推方程为

$$\begin{cases} f_k(s_k)=\max\limits_{u_k\in D_k(u_k)}[R_k(u_k)f_{k+1}(u_{k+1})] \qquad k=3,2,1 \\ C_k(u_k)\leqslant s_k-\sum\limits_{i=k+1}^{3}C_i(1) \\ f_4(s_4)=1 \end{cases}$$

其中 $C_k(u_k)\leqslant s_k-\sum\limits_{i=k+1}^{3}C_i(1)$ 表示首先要保证设备正常使用，即该设备的三个主要元件中的每个至少都要有一个，然后才能考虑加备件增加可靠性的措施，实际上增加可靠性的资金只有 4 万元。

1) 第三阶段($k=3$)

s_3 满足 $C_3(u_3)\leqslant s_3$，$C_3(1)\leqslant s_3\leqslant s_1-C_2(1)-C_1(1)$，所以 $2\leqslant s_3\leqslant 6$，迭代过程见表 7-8(b)。

表 7-8(b)　第三阶段

s_3	$D_3(s_3)$	s_4	$R_3(u_3)$	$R_3(u_3)\times f_4(s_4)$	$f_3(s_3)$	最优决策 u_3
2	1	0	0.5	0.5×1	0.5	1
3	1	1	0.5	0.5×1	0.5	1
4	1	2	0.5	0.5×1	0.7	2
	2	0	0.7	0.7×1		
5	1	3	0.5	0.5×1	0.9	3
	2	1	0.7	0.7×1		
	3	0	0.9	0.9×1		
6	1	4	0.5	0.5×1	0.9	3
	2	2	0.7	0.7×1		
	3	1	0.9	0.9×1		

北京大学出版社版权所有 禁止转载

2) 第二阶段($k=2$)

s_2 满足 $C_2(u_2)\leqslant s_2-C_3(1)$，$C_3(1)+C_2(1)\leqslant s_2\leqslant s_1-C_1(1)$，所以 $5\leqslant s_2\leqslant 9$，迭代过程见表 7-8(c)。

表 7-8(c)　第二阶段

s_2	$D_2(s_2)$	s_3	$R_2(u_2)$	$R_2(u_2)\times f_3(s_3)$	$f_2(s_2)$	最优决策 u_2
5	1	2	0.7	0.7×0.5	0.35	1
6	1	3	0.7	0.7×0.5	0.35	1
7	1	4	0.7	0.7×0.7	0.49	1
	2	2	0.8	0.8×0.5		
8	1	5	0.7	0.7×0.9	0.63	1
	2	3	0.8	0.8×0.5		
	3	2	0.9	0.9×0.5		
9	1	6	0.7	0.7×0.9	0.63	1
	2	4	0.8	0.8×0.7		
	3	3	0.9	0.9×0.5		

3) 第一阶段($k=1$)

$s_1=10$，迭代过程见表 7-8(d)。

表 7-8(d)　第一阶段

s_1	$D_1(s_1)$	s_2	$R_1(u_1)$	$R_1(u_1)\times f_2(s_2)$	$f_1(s_1)$	最优决策 u_1
10	1	9	0.6	0.6×0.63	0.504	2
	2	8	0.8	0.8×0.63		
	3	7	0.9	0.9×0.49		

北京大学出版社版权所有　禁止转载

最优的设计方案是 1 号元件并联 2 个；2 号元件并联 1 个；3 号元件并联 3 个，系统的可靠性为 0.504。

【例 7.12】 二维系统的可靠性问题。某个机器工作系统由 n 个部件串联而成，其中只要有一个部件失效，则整个系统不能正常工作，因此为了提高系统工作的可靠性，在设计时，每个主要部件上都装有备用元件，一旦某个主要部件失效，备用元件会自动投入系统工作，显然备用元件越多，系统工作可靠性越大，但是备用元件越多，系统的成本、质量、体积相应增大，工作精度降低，因此在上述限制条件下，应选择合理的备用元件数，使整个系统的工作可靠性最大。

解： 设第 $i(i=1,2,\cdots,n)$ 个部件上装有 u_i 个备用元件，正常工作的概率为 $p_i(u_i)$，则整个系统正常工作的可靠性为 $P=\prod_{i=1}^{n}p_i(u_i)$，装第 i 个部件的费用为 c_i，质量为 w_i。要求总费用不超过 c，总质量不超过 w，则线性规划数学模型为

$$\max P=\prod_{i=1}^{n}p_i(u_i)$$

$$\begin{cases}\sum_{i=1}^{n} c_i u_i \leqslant c \\ \sum_{i=1}^{n} w_i u_i \leqslant w \\ u_i \geqslant 0 \text{且为整数} \quad i=1,2,\cdots,n\end{cases}$$

应用动态规划方法求解的有关要素和动态规划基本方程如下。

按部件个数划分阶段，$k=1,2,\cdots,n$。

决策变量 u_k 表示部件 k 上的备用元件数。

状态变量 x_k 表示从第 k 个到第 n 个部件的总费用，

y_k 表示从第 k 个到第 n 个部件的总质量。

状态转移方程为

$$x_{k+1}=x_k-c_k u_k,\quad y_{k+1}=y_k-w_k u_k$$

允许决策集合为

$$D_k(x_k,y_k)=\left\{u_k \,\middle|\, 0\leqslant u_k \leqslant \min\left(\left[\frac{x_k}{c_k}\right],\left[\frac{y_k}{w_k}\right]\right),\text{且}u_k\text{为整数}\right\}$$

阶段指标函数为 $p_k(u_k)$，表示第 k 个部件的正常工作概率。

最优指标函数 $f_k(x_k,y_k)$ 表示由状态 x_k、y_k 出发，从部件 k 到部件 n 的系统工作最大可靠性，则动态规划基本方程为

$$\begin{cases}f_k(x_k,y_k)=\max\limits_{u_k\in D_k(x_k,y_k)}[p_k(u_k)f_{k+1}(x_{k+1},y_{k+1})] & k=n,n-1,\cdots,1 \\ f_{n+1}(x_{n+1},y_{n+1})=1\end{cases}$$

$f_1(c,w)$ 即为整个系统工作的最大可靠性。

北京大学出版社版权所有
禁止转载

7.4.4 设备更新问题

企业中经常会遇到一台设备应该使用多少年更新最合算的问题。一般来说，一台设备在比较新时，年运转量大、经济收入高、故障少、维修费用低，但随着使用年限的增加，年运转量减少因而收入减少、故障变多、维修费用增加。如果更新可提高年净收入，但是当年要支出一笔数额较大的购买费。

设备更新的一般提法为：在已知一台设备的效益函数 $r(t)$，维修费用函数 $u(t)$ 及更新费用函数 $c(t)$ 条件下，要求在 n 年中每年年初作出决策，是继续使用旧设备还是更换一台新设备，使 n 年总效益最大，参看例 7.13。

【例 7.13】某台新设备的年效益及年均维修费、更新净费用见表 7-9(a)。试确定今后 5 年内的更新策略，使总收益最大。

表 7-9(a) 新设备的年效益及年均维修费　　单位：万元

项目＼役龄	0	1	2	3	4	5
效益 $r_k(t)$	5	4.5	4	3.75	3	2.5
维修费 $u_k(t)$	0.5	1	1.5	2	2.5	3
更新费 $c_k(t)$	0.5	1.5	2.2	2.5	3	3.5

解：定义动态规划基本要素，建立动态规划求解方程。

以年限划分阶段 k， $k=1,2,3,4,5$。

决策变量 x_k： $x_k\begin{cases}K, & \text{在第 } k \text{ 年年初保留使用}\\ R, & \text{在第 } k \text{ 年年初更新}\end{cases}$

状态变量 s_k：第 k 年年初，设备已使用过的年数，称为役龄。

状态转移方程： $S_{k+1}\begin{cases}S_{k+1}, & x_k=K\\ 1, & x_k=R\end{cases}$

$r_k(t)$：在第 k 年设备已使用过 t 年(或役龄为 t 年)，再使用 1 年时的效益。

$u_k(t)$：在第 k 年设备已使用过 t 年(或役龄为 t 年)，再使用 1 年时的维修费用。

$c_k(t)$：在第 k 年卖掉一台役龄为 t 年的设备，买进一台新的设备的更新净费用，即 $c_k(t)$=买一部新机器的费用－卖一部 t 年役龄的旧机器的收益。

阶段指标函数

$$g_k(x_k)=\begin{cases}r_k(s_k)-u_k(s_k), & x_k=K\\ r_k(0)-u_k(0)-c_k(s_k), & x_k=R\end{cases}$$

最优指标函数 $f_k(s_k)$：第 k 年年初，使用一台已用了 s_k 年的设备，到第 5 年年末的最大效益，有

$$f_k(s_k)=\max\begin{cases}r_k(s_k)-u_k(s_k)+f_{k+1}(s_{k+1}), & x_k=K\\ r_k(0)-u_k(0)-c_k(s_k)+f_{k+1}(1), & x_k=R\end{cases}$$

建立逆序递推基本方程为

$$\begin{cases}f_k(s_k)=\max\limits_{x_k=K,R}\{g_k(x_k)+f_{k+1}(s_{k+1})\}, & k=5,4,3,2,1\\ f_6(s_6)=0\end{cases}$$

下面用逆序法求解。

1) 第五阶段($k=5$)

此时 s_5 的所有可能取值为 1、2、3、4，见表 7-9(b)。

$$f_5(s_5)=\max_{x_5=K,R}\{g_5(x_5)+f_6(s_6)\}=\max\begin{cases}r_5(s_5)-u_5(s_5), & x_5=K\\ r_5(0)-u_5(0)-c_5(s_5), & x_5=R\end{cases}$$

表 7-9(b)　第五阶段

s_5	$D_5(s_5)$	s_6	$g_5(x_5)$	$g_5(x_5)+f_6(s_6)$	$f_5(s_5)$	最优决策 x_5
1	K	2	4.5−1	4.5−1	3.5	K
	R	1	5−0.5−1.5	5−0.5−1.5		
2	K	3	4−1.5	4−1.5	2.5	K
	R	1	5−0.5−2.2	5−0.5−2.2		
3	K	4	3.75−2	3.75−2	2	R
	R	1	5−0.5−2.5	5−0.5−2.5		
4	K	5	3−2.5	3−2.5	1.5	R
	R	1	5−0.5−3	5−0.5−3		

2) 第四阶段($k=4$)

此时， s_4 的所有可能取值为 1、2、3，见表 7-9(c)。

北京大学出版社版权所有 禁止转载

表 7-9(c)　第四阶段

s_4	$D_4(s_4)$	s_5	$g_4(x_4)$	$g_4(x_4)+f_5(s_5)$	$f_4(s_4)$	最优决策 x_4
1	K R	2 1	4.5−1 5−0.5−1.5	4.5−1+2.5 5−0.5−1.5+3.5	6.5	R
2	K R	3 1	4−1.5 5−0.5−2.2	4−1.5+2 5−0.5−2.2+3.5	5.8	R
3	K R	4 1	3.75−2 5−0.5−2.5	3.75−2+1.5 5−0.5−2.5+3.5	5.5	R

3) 第三阶段($k=3$)

此时，s_3 的所有可能取值为 1、2，见表 7-9(d)。

表 7-9(d)　第三阶段

s_3	$D_3(s_3)$	s_4	$g_3(x_3)$	$g_3(x_3)+f_4(s_4)$	$f_3(s_3)$	最优决策 x_3
1	K R	2 1	4.5−1 5−0.5−1.5	4.5−1+5.8 5−0.5−1.5+6.5	9.5	R
2	K R	3 1	4−1.5 5−0.5−2.2	4−1.5+5.5 5−0.5−2.2+6.5	8.8	R

4) 第二阶段($k=2$)

此时，s_2 的取值只能为 1，计算过程见表 7-9(e)。

表 7-9(e)　第二阶段

s_2	$D_2(s_2)$	s_3	$g_2(x_2)$	$g_2(x_2)+f_3(s_3)$	$f_2(s_2)$	最优决策 x_2
1	K R	2 1	4.5−1 5−0.5−1.5	4.5−1+8.8 5−0.5−1.5+9.5	12.5	R

5) 第一阶段($k=1$)

此时，s_1 的取值只能是 0，所以

$$f_1(0)=\max\begin{Bmatrix} 5-0.5+12.5 \\ 5-0.5-0.5+12.5 \end{Bmatrix}=17$$

此时最优决策 $x_1=K$，按照上述计算递推回去，故本题最优策略为$\{K，R，R，R，K\}$，即第一年年初购买的设备到第二、三、四年年初各更新一次，用到第五年年末，其总效益为 17 万元。

7.4.5　背包问题

【例 7.14】 某人携带背包上山，其可携带物品的质量限度为 a 千克，现有 n 种物品可供选择，设第 i 种物品的单件质量为 a_i 千克，其在上山过程中的价值是携带数量 x_i 的函数 $c_i(x_i)$，问应如何安排携带各种物品的数量，使总价值最大。

北京大学出版社版权所有
禁止转载

解：背包问题的数学模型为

$$\max z = c_1(x_1) + c_2(x_2) + \cdots + c_n(x_n)$$

$$\begin{cases} a_1x_1 + a_2x_2 + \cdots + a_nx_n \leqslant a \\ x_i \geqslant 0\text{且为整数}(i = 1, 2, \cdots, n) \end{cases}$$

下面用动态规划方法求解。

按照装入物品的种类划分阶段，$k = 1, 2, \cdots, n$。

状态变量为s_k，表示分配给第k个至第n种物品的总重量。

决策变量x_k表示装入第k种物品的件数，允许决策集合为

$$D_k(s_k) = \left\{ x_k \middle| 0 \leqslant x_k \leqslant \left[\frac{s_k}{a_k}\right],\ x_k\text{为整数} \right\}$$

其中$\left[\frac{s_k}{a_k}\right]$表示不超过$\frac{s_k}{a_k}$的最大整数。

状态转移方程：$s_{k+1} = s_k - a_kx_k$，其中$s_1 = a$。

阶段指标函数$c_k(x_k)$：表示第k阶段装入第k种商品x_k件时的价值。

最优指标函数$f_k(s_k)$：表示第k阶段装入物品总重量为s_k时的最大价值，动态规划递推方程为

$$\begin{cases} f_k(s_k) = \max\limits_{x_k = 0, 1, \cdots, \left[\frac{s_k}{a_k}\right]} [c_k(x_k) + f_{k+1}(s_{k+1})] \qquad k = n, n-1, \cdots, 1 \\ f_{n+1}(s_{n+1}) = 0 \end{cases}$$

由后向前逐段递推，$f_1(a)$即为所求问题的最大收益。

【例 7.15】 设有一辆载重量为 10 吨的卡车，用以装载 3 种货物，每种货物的单件质量及单件价值见表 7-10(a)。问各种货物应装多少件，才能既不超过总质量(以吨为单位计)又使总价值最大。

北京大学出版社版权所有 禁止转载

表 7-10(a)　每种货物的单件质量及单件价值

货　物	1	2	3
单件质量	3	4	5
单件价值	4	5	6

解：设x_j表示装载第j种货物的件数，$j = 1, 2, 3$，则问题可归结为

$$\max z = 4x_1 + 5x_2 + 6x_3$$

$$\begin{cases} 3x_1 + 4x_2 + 5x_3 \leqslant 10 \\ x_i \geqslant 0\text{且为整数} \quad (i = 1, 2, 3) \end{cases}$$

这是一个整数规划问题，下面用动态规划方法来解。

设阶段变量$k = 1, 2, 3$共分 3 个阶段，状态变量s_k表示从第k阶段至第 3 阶段可供装载的总质量；决策变量x_k表示第k种货物装载的件数，且x_k取整数，允许决策集合为

$$D_k(s_k) = \left\{ x_k \middle| 0 \leqslant x_k \leqslant \left[\frac{s_k}{a_k}\right], x_k\text{为整数} \right\}$$

则状态转移方程为 $s_{k+1}=s_k-a_kx_k$，其中 a_k 表示第 k 种货物的单件质量；阶段指标函数 $c_k(x_k)$ 表示第 k 阶段装入第 k 种商品 x_k 件时的价值。

最优指标函数 $f_k(s_k)$ 表示第 k 阶段装入物品总重量为 s_k 时的最大价值，动态规划基本方程为

$$\begin{cases} f_k(s_k)=\max\limits_{x_k=0,1,\cdots,\left[\frac{s_k}{a_k}\right]}[c_k(x_k)+f_{k+1}(s_{k+1})] & k=3,2,1 \\ f_4(s_4)=0 \end{cases}$$

采用动态规划的逆序解法，求解上面动态规划。

1) 第三阶段($k=3$)

$$f_3(s_3)=\max_{0\leqslant x_3\leqslant\left[\frac{s_3}{5}\right]}\{6x_3+f_4(s_4)\}=\max_{0\leqslant x_3\leqslant\left[\frac{s_3}{5}\right]}\{6x_3\}=6\left[\frac{s_3}{5}\right],\quad x_3=\left[\frac{s_3}{5}\right]$$

其中 $[x]$ 表示不超过 x 的最大整数。因此当 $s_3=0,1,2,3,4$ 时，$x_3=0$；当 $s_3=5,6,7,8,9$ 时，x_3 可取 0 或 1；当 $s_3=10$，x_3 可取 0、1 或 2，由此确定 $f_3(s_3)$，现将有关数据填入表 7-10(b)中。

表 7-10(b)　第三阶段

s_3	$D_3(s_3)$	s_4	$6x_3$	$6x_3+f_4(s_4)$	$f_3(s_3)$	最优决策 x_3
0	0	0	0	0	0	0
1	0	1	0	0	0	0
2	0	2	0	0	0	0
3	0	3	0	0	0	0
4	0	4	0	0	0	0
5	0 1	5 0	0 6	0 6	6	1
6	0 1	6 1	0 6	0 6	6	1
7	0 1	7 2	0 6	0 6	6	1
8	0 1	8 3	0 6	0 6	6	1
9	0 1	9 4	0 6	0 6	6	1
10	0 1 2	10 5 0	0 6 12	0 6 12	12	2

北京大学出版社版权所有　禁止转载

2) 第二阶段($k=2$)

$$f_2(s_2)=\max_{0\leqslant x_2\leqslant\left[\frac{s_2}{4}\right]}\{5x_2+f_3(s_3)\}=\max_{0\leqslant x_2\leqslant\left[\frac{s_2}{4}\right]}\{5x_2+f_3(s_2-4x_2)\}$$

当 $s_2=0,1,2,3$ 时，$x_2=0$；当 $s_2=4,5,6,7$ 时，x_2 可取 0 或 1；当 $s_2=8,9,10$ 时，x_2 可取 0、1 或 2，由此确定 $f_2(s_2)$。现将有关数据填入表 7-10(c)中。

表 7-10(c)　第二阶段

s_2	$D_2(s_2)$	s_3	$5x_2$	$5x_2+f_3(s_3)$	$f_2(s_2)$	最优决策 x_2
0	0	0	0	0+0	0	0
1	0	1	0	0+0	0	0
2	0	2	0	0+0	0	0
3	0	3	0	0+0	0	0
4	0 1	4 0	0 5	0+0 5+0	5	1
5	0 1	5 1	0 5	0+6 5+0	6	0
6	0 1	6 2	0 5	0+6 5+0	6	0
7	0 1	7 3	0 5	0+6 5+0	6	0
8	0 1 2	8 4 0	0 5 10	0+6 5+6 10+0	11	1
9	0 1 2	9 5 1	0 5 10	0+6 5+6 10+0	11	1
10	0 1 2	10 6 2	0 5 10	0+12 5+6 10+0	12	0

2) 第一阶段($k=1$)

$$f_1(s_1)=\max_{0\leqslant x_1\leqslant\left[\frac{s_1}{3}\right]}\{4x_1+f_2(s_2)\}=\max_{0\leqslant x_1\leqslant\left[\frac{s_1}{3}\right]}\{4x_1+f_2(s_1-3x_1)\}$$

由于 $s_1=10$，所以 x_1 可取 0、1、2 或 3，由此确定 $f_1(s_1)$。现将有关数据填入表 7-10(d)中。

表 7-10(d)　第一阶段

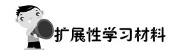
s_1	$D_1(s_1)$	s_2	$4x_1$	$4x_1+f_2(s_2)$	$f_1(s_1)$	最优决策 x_1
10	0 1 2 3	10 7 4 1	0 4 8 12	0+12 4+6 8+5 12+0	13	2

由表 7-10(d)可知，$f_1(s_1)$ 取得最大值 13，最优解为 $x_1=2,\quad x_2=1,\quad x_3=0$，即第一种货物装 2 件，第二种货物装 1 件，第三种货物不装，可使总价值最大。

从上面应用实例可知，一般来讲只要多阶段决策问题满足无后效性，且整个过程具有递推性，这样的多阶段决策问题都可用动态规划方法来处理。运用动态规划解决实际问题易于确定全局最优解，但是动态规划方法也存在不足之处：一是没有统一的标准模型可供使用，建模时需要根据问题性质及数学特点加以解决，这就需要灵活的技巧及大量的实践；二是应用局限性，动态规划模型中的状态变量必须满足无后效性，而许多实际问题往往不

北京大学出版社版权所有 禁止转载

能满足这一条件；三是存在“维数障碍”，即当变量个数(维数)太大时，问题虽然可以用动态规划方法来描述，但由于计算机内存容量的限制而难以求解，一般超过三维(含三维)的问题通常不采用动态规划方法来求解。



扩展性学习材料

动态规划在语音识别中的应用

在计算机算法设计方法中，动态规划技术是比较基本，但又比较抽象，难于理解的一种，它建立在最优原则的基础上，采用动态规划方法，可以优雅而高效地解决许多用贪心技术或分治技术无法解决的问题。因此，动态规划技术越来越成为解决许多重要的应用问题的关键技术。如用动态规划解决 0-1 背包问题、图像数据压缩、矩阵连乘、有向图最短路径、无交叉子集、元件折叠，以及最长公共子序列等应用问题。

在过去的语音识别技术的发展中，最具决定性意义的除了声学建模单元的选择外就是搜索技术的发展。在最初的识别系统中，大都是采用动态规划的方法，基于动态规划的搜索技术被广泛地应用于很多很成熟的语音识别系统中，最初的是数字识别、孤立词的模板匹配和孤立词的马尔科夫模型识别，其中应用动态规划技术的动态时间伸缩算法 DTW 取得了很大成功，当词汇表较小以及各个词条不易于混淆时，DTW 可以有效地解决孤立词识别时说话速度不均匀的难题，从而自 20 世纪 60 年代末期掀起了语音识别研究的热潮。

动态时间伸缩算法(Dynamic Time Wrapping，DTW)，是日本学者板仓(Itakura)将动态规划技术应用于解决孤立词识别时的说话速度不均匀的难题，提出的把时间规整和距离测度计算结合起来的一种非线性归整技术。动态规划还用于解决字符串搜索问题、手写字符识别问题、网络的无交叉布线、电路元件折叠和旅行商问题等各种实际应用问题。

北京大学出版社版权所有 禁止转载

本 章 小 结

动态规划是研究多阶段决策的最优化方法，多阶段决策问题含有一个描述过程时序或空间演变的阶段变量，将复杂问题划分成若干阶段，根据动态规划“最优性原理”，逐段解决而最终实现全局最优。本章阐述了动态规划基本要素、最优化原理和基本方程，通过资源分配和生产与存储等问题，示范了动态规划解决多阶段决策问题的过程，将动态规划应用到经济、管理、工业生产、工程技术等领域。



关键术语(中英文)

动态规划(Dynamic Programming)
阶段变量(Stage Variable)
状态变量(State Variable)
决策变量(Decision Variable)
策略(Policy)
阶段指标(Stage Indicator)
指标函数(Indicator Function)
最优性原理(Principle of Optimality)
生产存储问题(Production and Inventory Problem)
基本方程(Basic Equation)
逆序解法(Inverse Order Method)
顺序解法(Order Method)

多阶段决策问题(Multistage Decision Problem)　状态转移方程(State Transition Equation)
设备更新问题(Equipment Replacement Problem)　背包问题(Knapsacks Problem)



知识链接

动态规划——贝尔曼

贝尔曼，美国数学家，美国全国科学院院士，动态规划的创始人。1920 年 8 月 26 日生于美国纽约，1984 年 3 月 19 日逝世。1941 年在布鲁克林学院毕业，获理学学士学位，1943 年在威斯康星大学获理学硕士学位，1946 年在普林斯顿大学获博士学位。1946—1948 年在普林斯顿大学任助理教授，1948—1952 年在斯坦福大学任副教授，1953—1956 年在美国兰德公司任研究员，1956 年后在南加利福尼亚大学任数学教授、电气工程教授和医学教授。

贝尔曼因提出动态规划而获美国数学会和美国工程数学与应用数学会联合颁发的第一届维纳奖金(1970)，卡内基－梅隆大学颁发的第一届迪克森奖金(1970)，美国管理科学研究会和美国运筹学会联合颁发的诺伊曼理论奖金(1976)。1977 年贝尔曼当选为美国艺术与科学研究院院士和美国工程科学院院士。

贝尔曼因在研究多段决策过程中提出动态规划而闻名于世。1957 年他的专著《动态规划》出版后，被迅速译成俄文、日文、德文和法文，对控制理论界和数学界有深远影响。贝尔曼还把不变嵌入原理应用于理论物理和数学分析方面，把两点边值问题化为初值问题，简化了问题的分析和求解过程。1955 年后贝尔曼开始研究算法、计算机仿真和人工智能，把建模与仿真等数学方法应用到工程、经济、社会和医学等方面，取得许多成就。贝尔曼对稳定性的矩阵理论、时滞系统、自适应控制过程、分岔理论、微分和积分不等式等方面都有过贡献。贝尔曼曾是《数学分析与应用杂志》及《数学生物科学杂志》的主编，《科学与工程中的数学》丛书的主编。出版 30 本著作和 7 本专著，发表了 600 多篇研究论文。

北京大学出版社版权所有 禁止转载



习题 7

7.1 填空题

(1) 动态规划是解决__________决策过程中最优化问题的一种方法。

(2) 动态规划中的 Bellman 最优性原理是______________________________。

(3) 对于多阶段决策问题来说，状态不仅要描述过程的具体特征，而且一个根本的要求是必须满足__________。

(4) 用动态规划方法求解背包问题时，将装载的__________作为状态。

(5) 在生产与存储问题中，状态变量为__________，决策变量是__________。

7.2 判断题

(1) 动态规划的基本方程是将一个多阶段的决策问题转化为一系列具有递推关系的单阶段决策问题。（　）

(2) 最优指标函数 $f_k(s_k)$ 表示第 k 阶段状态为 s_k 时到下一阶段的最优指标值。（　）

(3) 动态规划数学模型由阶段、状态、决策与策略、状态转移方程及指标函数这 5 个要素组成。（　）

(4) 连乘形式的递推方程的边界条件等于 1，连和形式的递推方程的边界条件等于 0。（　）

(5) 动态规划是一种将问题分解为更小的、相似的子问题，并存储子问题的解而避免计算重复的子问题，以解决最优化问题的算法策略。 (　　)

7.3 设某公司计划在 1～4 月从事某种商品经营。已知仓库最多可存储 600 件这种商品，已知 1 月初存货 200 件，根据预测知 1～4 月的单位购货成本及销售价格见表 7-11，每月只能销售本月月初的库存，当月进货供以后各月销售。假设 4 月底库存为零，问如何安排进货量和销售量，使该公司 4 个月获得利润最大(仅要求列出动态规划的基本要素和基本方程，不求解)？

表 7-11　1～4 月的单位购货成本及销售价格

月　份	购货成本 C	销售价格 P
1	40	45
2	38	42
3	40	39
4	42	44

7.4 有一部货车沿着公路给 4 个零售店卸下 6 箱货物。如果各零售店出售该货物所得利润见表 7-12，试求在各零售店卸下多少箱货物，能使获得的总利润最大？

表 7-12　各零售店出售货物的利润表

零售店 箱　数	1	2	3	4
0	0	0	0	0
1	4	2	3	4
2	6	4	5	5
3	7	6	7	6
4	7	8	8	6
5	7	9	8	6
6	7	10	8	6

北京大学出版社版权所有　禁止转载

7.5 某企业有某种高效率设备 3 台，拟分配给所属甲、乙、丙车间，各车间得到设备后，获利情况见表 7-13，试给出最优分配方案。

表 7-13　收益表

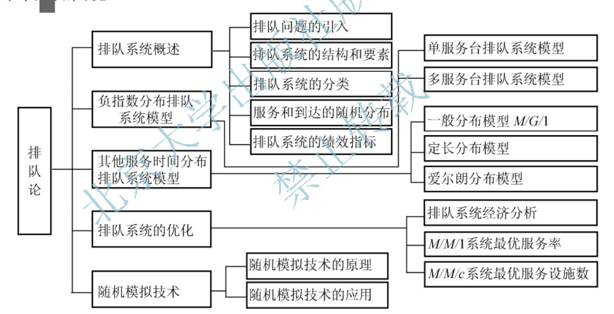
工　厂 设备台数	甲	乙	丙
0	0	0	0
1	3	5	4
2	7	10	6
3	9	11	11

7.6 某工程在一年内拟进行 A、B、C 三种新产品试制，由于资金不足，估计年内这 3 种新产品研制不成功的概率分别为 0.40、0.60、0.80，因而都研制不成功的概率为这 3 个概率的乘积(0.192)。为了保证三种新产品研制的成功率，决定增拨 2 万元的研制经费，并要资金集中使用，以万元为单位进行分配，其增拨研制费与新产品不成功的概率见表 7-14。试问如何分配费用，使得三种新产品研制都不成功的概率为最小？

第8章 排 队 论

北京大学出版社版权所有 禁止转载

【本章知识架构】



【本章教学目标与要求】

- 熟悉排队系统的基本结构与三大要素；掌握排队论的常用术语和记号；熟练识别康道尔排队系统的分类规则；熟悉排队系统常见的输入、输出分布。
- 熟练掌握负指数分布排队系统的绩效指标的计算方法，并能进行不同系统间的比较。
- 了解一般分布时间的排队系统的绩效指标的计算方法。
- 掌握对简单排队系统的优化分析方法。
- 了解随机服务系统的模拟思路，能对简单排队系统进行模拟分析。



运储物流客户服务中心

客户服务是整个物流体系设计和运作的必要部分，差异化的客户服务能给企业带来独特的竞争优势，在短期内是竞争对手难以模仿的。物流的本质就是服务，运储物流公司跑步进入了“以客户为中心”的市场竞争时代，成功地打造了在线客户服务系统。它作为企业的核心服务系统，是集企业形象宣传、信息查询、业务受理、客户投诉及建议、客户回访与主动呼出为一体的企业级多功能呼叫中心平台系统。该物流客户服务系统和原有的物流管理系统连接，实时地将客户的数据导入呼叫中心数据库。

在线服务系统的设计除了要融合现代物流的先进理念和信息、物流技术外，还必须考察客户的呼叫次数、平均服务时间，以及服务台的人员配备等要素，应用排队理论来优化系统，有效地实现了产品与信息从原料到最终消费者之间的增值服务，最终提高公司的服务效率，提升客户的满意度和忠诚度，为企业自身带来更多的效益。

资料来源： http://www.ctiforum.com.

北京大学出版社版权所有 禁止转载

排队论(Queuing Theory)，或称随机服务系统理论，是运筹学重要的分支学科，着重研究服务系统中排队现象的随机规律，解决相应排队系统的最优设计和最优控制问题。排队论的产生与发展来自实际的需要，应用非常广泛，尤其是在通信、交通、存储、生产管理等资源共享的随机服务领域。

8.1 排队系统概述

排队是指在服务系统中要求服务的“顾客”所形成的等待线。鉴于服务系统中顾客的到达及服务时间和次序都是随机的，排队论通过对服务对象到来及服务时间的统计研究，得出这些数量指标(等待时间、排队长度、忙期长短等)的统计规律，然后根据这些规律来改进服务系统的结构或重新组织被服务对象，使得服务系统既能满足服务对象的需要，又能使服务机构的费用最经济或某些指标最优。

8.1.1 排队问题的引入

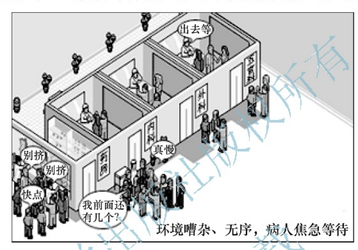
排队与日常生活密切相关，在火车站购票、奥运会安全检查、中国移动营业窗口办理手机业务时经常排队。尽管我们对此已经习以为常，但经常为排队而抱怨。消耗在排队上的时间是影响生活质量和国家经济质量的一个主要因素。在商品短缺的时代，人们为购买日常必需品经常不得不排极长的队列；即使在今日的美国，据统计人们每年消耗在排队中的时间达370亿小时，如果将这些时间用于生产，几乎是一年中2 000万人的有效工作时间。

然而这些令人吃惊的数据还没有反映出过长的等待所造成的全部影响，除此之外，其他类型的排队也会导致严重的低效率，例如生产设备维修造成生产损失，运输工具因为等待卸货而耽误后面的运输，飞机等待起飞或降落而打乱后面的飞行计划。排队论研究所有这些情况下的等待，为便于分析，常常使用排队模型来表示各种类型的排队系统，每一个排队模型都描述了对应的排队系统的工作原理，首先看几个典型的排队问题。

【例 8.1】 医院挂号排队。

卫生部近日频频下发文件，以推进医疗改革，继《医院投诉管理办法(征求意见稿)》之后，又下发了《关于在公立医院施行预约诊疗服务工作的意见(征求意见稿)》。要求从 2009 年 10 月开始，所有三级医院都要开展预约诊疗服务，在预约挂号时要推广采取实名制方式。

也许有人不太理解挂号实名制的含义，认为只是挂一个号，不是什么大事情，但如果到基层医院挂过号排过队，就容易明白拥挤和排队所造成的低效率。由于专家资源稀缺，导致其能提供的医疗服务非常有限，虽然专家门诊的挂号费用贵于普通门诊，患者仍会慕名前来挂专家门诊，甚至要彻夜排队挂号。由于匿名，就容易出现特权者插队的情况，这就导致患者即便挂到了号，也仍需等待很长的时间才能获得相应的专家门诊服务。



北京大学出版社版权所有
禁止转载

【例 8.2】 银行存取款服务。

银行作为过去的大型的国有企业，在经济全球化的进程中也已经走向了市场的竞争，国外的很多大银行也加快了中国市场的开发。如何能够为顾客提供更为舒适、个性化的服务是国内的银行迫在眉睫的问题。多年来“一米线”的服务已满足不了人们的需求。站立等候已经过时，提供舒适的环境已成竞争的重要手段，许多银行营业厅不仅有了沙发、报纸，还有纯净水、糖果和咖啡。

【例 8.3】 高速公路收费。

我国采用多渠道融资、分段修建的方式建设高速公路，各路段由其业主公司自己管理、设站收费，使得高速公路网内收费站数量过多，严重降低了车辆通行能力和高速公路的使用效率。高速公路收费站是公路交通流的瓶颈，数据调查研究表明车辆在高速公路上的延误时间中有 36%是由于停车收费原因造成的，特别是在“黄金周”期间，高速公路上等候的旅游车辆几乎是一步一挪前行，许多高速路段从某个收费站就开始排起长龙，逐渐绵延至七八千米，形成节日里一道特殊的景观。

【例 8.4】 运输服务中的排队系统，见表 8-1。

表 8-1 运输服务中的排队系统

系统类型	顾　　客	服　务　台
港口装卸区	轮船	装卸人员
停车场	汽车	停车可控关键

续表

系统类型	顾 客	服务台
等待起飞的飞机	飞机	服务台
到达物流配送中心的商品	货物配送	物流配送中心
消防部门	火灾	急救车

行为科学家发现：无序排队是影响客户流失的一条主要原因。研究结果表明：等候超过十分钟，情绪开始急躁；超过二十分钟，情绪表现厌烦；超过四十分钟，常因恼火而离去，而其中如出现“加塞”“插队”现象，情况还将更加糟糕。服务窗口是形成银行、电信、航空、医院等企业的公众形象的重要因素，如何解决长久以来枯燥的排队问题，创造一个轻松的个性化的窗口环境，就显得非常重要。

排队论研究的内容有 3 个方面：统计推断，根据资料建立模型；系统的性态，即和排队有关的数量指标的概率规律性；系统的优化问题，其目的是正确设计和有效运行各个服务系统，使之以最经济的服务机构满足顾客的需要。

8.1.2 排队系统的结构和要素

排队论里把要求服务的对象统称为“顾客”，而把提供服务的人或机构称为“服务台”，不同的顾客与服务台组成了各式各样的服务系统，顾客为了得到某种服务而到达系统，若不能立即获得服务而又允许排队等待，则加入等待队伍。一个典型的排队系统如图 8.1 所示。顾客一个挨着一个接受某种服务。如果一位到达的顾客不能立即接受服务，这个顾客就加入到一个队列中(这个队列不包括正在接受服务的顾客，由排队规则决定)。在服务机构中有一个或多个服务台提供服务。每一位顾客都被单独地安排在其中一个服务台服务，完成服务后自行离开。

北京大学出版社版权所有 禁止转载

服务系统的顾客，可以是人，也可以是一些机器设备。然而无论哪种服务系统，都可表示为图 8.1 所示的排队模式。一般排队系统由输入过程、排队规则和服务机构三要素组成。

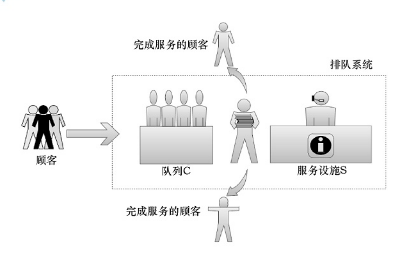


图 8.1 排队系统的结构

1. 输入过程

输入过程考察的是顾客到达服务系统的规律。它可以用一定时间内顾客的到达数或前

后两个顾客相继到达的间隔时间来描述，一般分为确定型和随机型两种。例如，在生产线上加工的零件按规定的间隔时间依次到达加工地点，定期运行的班车、班机等都属于确定型输入。随机型的输入是指在时间 t 内顾客到达数 $N(t)$ 服从一定的随机分布，下面将详细探讨。在排队论中，讨论的输入过程主要是随机型的。

2. 排队规则

排队规则有以下 3 种。

1) 损失制

顾客到达时，若所有服务台都正被占用，在这种情况下顾客就随即离去，永不再来。如进入电话交换台的“呼叫”，或进入空防区的“敌机”，就可看做是损失制。

2) 等待制

顾客到达时，若所有服务台均被占用，顾客不离去，而是排队等待服务。其排队的方式和服务的次序可采用各种各样的规则。

(1) 单列排队。到达顾客排成一列，按规定次序接受服务。

(2) 多列排队。到达顾客排成多列，接受服务。各列之间，顾客可相互转移，有的顾客因等待时间过长而中途退出，有的顾客不中途退出，如高速公路上的汽车流，必须坚持到服务为止。这里只讨论各队列间不能互相转移也不能中途退出的情形。

(3) 先到先服务。按到达次序先后，先到的顾客先服务，后到的顾客后服务。这种自然的服务次序是最常见的规则，这里主要讨论这种排队规则。

(4) 后到先服务。按到达次序，最后到达的顾客首先接受服务。如将堆入仓库的钢板看作顾客，使用时陆续取出作为服务，则一般是取用放在最上面的钢板；又如计算机堆栈里的数据，总是首先弹出最后进入堆栈的数据。

(5) 优先权服务。按到达顾客的不同身份而优先服务某些顾客，如邮政中的特别邮件，医院对危重病人要优先治疗等。

(6) 随机服务。当服务台得空时，在排队的顾客中随机地选取一名服务，即每一顾客被选到的概率相同。

3) 混合制服务

根据系统情况，顾客到达时，既可能离去也可能排队，即损失制和等待制兼而有之的排队规则。例如，就加油站而言，由于场地限制，到达的车辆可能进入加油站排队等待加油，也可能由于加油站已排满车辆而自动离去。

3. 服务机构

服务机构主要考虑服务台的数量、服务方式和服务时间的规律。

(1) 服务机构的数量是一个或多个服务台。多个服务台可以是平行排列的，也可以是串联排列的，还可能是串、并行混合排列。图 8.2(a)是单队-单服务台，(b)是多队多-多服务台，(c)是单队-并列多服务台，(d)是单队-串列多服务台，(e)是混合的情形。

(2) 服务方式。可以是对单个顾客服务，如理发馆、零件加工、机器维修等，也可以是成批的服务，如餐馆、旅客列车、客运飞机等。这里只研究单个服务方式。

(3) 服务时间一般也分成确定型和随机型两种。例如，自动冲洗汽车的装置对每辆汽车冲洗(服务)时间是相同的，因而是确定型的；而随机型服务时间 V 则服从一定的随机分

北京大学出版社版权所有 禁止转载

布，大致有定长分布、负指数分布、Erlang 分布、一般独立分布等。但值得注意的是，在一个排队系统中，输入过程和服务时间的分布可以是不同的。

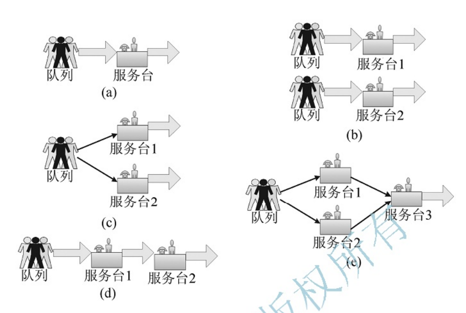


图 8.2 各类服务系统

北京大学出版社版权所有 禁止转载

8.1.3 排队系统的分类

按照排队系统 3 个组成部分的特征的各种可能情形来分类，可以给出许多排队系统的类型。为了方便对排队系统的描述，康道尔(D G Kendall)在 1953 年提出一种根据排队系统的 3 个基本特征(相继顾客到达系统的间隔时间分布、服务时间的分布和服务台数目)对排队模型分类的表示方法，即 Kendall 记号，一般形式为

$$X/Y/Z$$

其中：X 处填写相继到达间隔时间的分布；Y 处填写服务时间分布；Z 处填写并列的服务台数目。

顾客到达时间间隔和服务时间的分布，除了最常用的负指数分布之外，还有其他一些分布，其含义和符号列举如下：M ——负指数分布(M 是 Markov 的字头，因为负指数分布具有 Markov 性)；D ——确定性(Deterministic)分布；E_k —— k 阶爱尔朗(Erlang)分布；G ——一般(General)服务时间分布。

例如，$M/M/1$ 表示顾客相继到达的间隔时间为负指数分布、服务时间为负指数分布和单个服务台的模型。$D/M/c$ 表示顾客按确定间隔时间到达、服务时间为负指数分布和 c 个服务台的模型。

后来在 1971 年的一次关于排队论符号标准化会议上决定，将 Kendall 记号扩充为

$$X/Y/Z/A/B/C$$

其中前三项的含义不变，而 A 表示系统容量限制(容纳的顾客数量)；B 表示顾客源的数目；C 表示服务规则，如先到先服务或后到先服务等，并约定如略去后三项，就表示是 $X/Y/Z/\infty/\infty/FCFS$ 这种类别，在本章中约定只讨论先到先服务这种情况，所以略去第六项。

8.1.4 服务和到达的随机分布

排队论主要涉及的输入过程和服务过程是随机的，描述它们的数学模型自然离不开概

率分布，下面简要介绍排队论常用的概率分布。

1. 泊松过程

泊松(Poisson)过程又称为泊松流，是排队论中经常用来描述顾客到达规律的特殊随机过程，与概率论中的泊松分布和负指数分布有密切的联系。

定义 8.1 泊松过程：设 $N(t)$ 为时间 $[0, t]$ 内到达系统的顾客数，如果满足以下 3 个条件。

(1) 独立性。即在任意两个不相重叠的时间区间内到达的顾客数是相互独立的，称这个性质为无后效性。

(2) 平稳性。对于充分小的 Δt，在时间区间 $[t, t+\Delta t]$ 内仅有一个顾客到达的概率与 t 无关，与 Δt 成正比，即 $P_1(t, t+\Delta t) = \lambda\Delta t + 0(\Delta t)$。

(3) 普通性。在时间区间 $[t, t+\Delta t]$ 内有多于一个以上的顾客到达的概率为无穷小量，即

$$\sum_{n=2}^{\infty} P_n(t, t+\Delta t) = 0(\Delta t)$$

则称 $\{N(t), t \geqslant 0\}$ 为泊松过程，那么在时间 $[0, t]$ 内到达系统的顾客数 $N(t)$ 的概率为

$$P_n(t) = P\{N(t) = n\} = \frac{(\lambda t)^n}{n!}\mathrm{e}^{-\lambda t} \quad n = 0, 1, 2, \cdots \tag{8-1}$$

公式(8-1)的形式与概率论中的泊松分布形式一样，只是在参数 λ 里增加时间参数 t 变成了 λt。泊松过程是在一定时间 t 内顾客到达系统的人数服从参数为 λt 的泊松分布，记为 $N(t) \sim P(\lambda t)$。也可以这样说，在单位时间内到达的顾客数 N 是服从泊松分布的随机变量，即服从

$$P\{N = n\} = \frac{(\lambda)^n}{n!}\mathrm{e}^{-\lambda} \quad n = 0, 1, 2, \cdots$$

现实中有许多现象服从泊松分布，如单位时间内通过某路口的车辆数(交通流)，单位时间内某交换台接到的呼叫次数(电话流)，某超市到达的顾客人数(顾客流)等。泊松分布其均值和方差分别为 $\mathrm{E}(N) = \lambda$，$\mathrm{D}(N) = \lambda$。

2. 负指数分布

负指数(Negative Exponential)分布也简称为指数分布，是排队论分析中用得最多的极为重要的概率分布，通常用来描述顾客到达的间隔时间和对顾客服务时间的分布。现实生活和实际活动中大多数排队问题符合这种分布规律。

定义 8.2 指数分布：设随机变量 T 的分布密度函数

$$f_T(t) = \begin{cases} \lambda \mathrm{e}^{-\lambda t}, & t \geqslant 0 \\ 0, & t < 0 \end{cases}$$

则称 T 为服从参数 λ 的负指数分布($\lambda > 0$)，记为 $T \sim E(\lambda)$，如图 8.3 所示。它的分布函数为

$$F_T(t) = P(T \leqslant t) = \begin{cases} 1-\mathrm{e}^{-\lambda t}, & t \geqslant 0 \\ 0, & t < 0 \end{cases}$$

其均值 $\mathrm{E}(T) = 1/\lambda$，方差 $\mathrm{D}(T) = 1/\lambda^2$。

北京大学出版社版权所有 禁止转载

负指数分布最大的特点是“无记忆性”，即马尔科夫(Markov)性，这为分析随机过程带来了极大的方便。这样可以在任意时刻观察系统，而保持其概率分布规律不变，这是负指数分布在排队论(以及随机过程)分析中用得较广的原因之一。

根据泊松过程的公式(8-1)，在$[0, t]$时间内没有顾客到达系统的概率为$P_0(t) = \mathrm{e}^{-\lambda t}$，那么至少有1个顾客到达的概率应是$1 - P_0(t) = 1 - \mathrm{e}^{-\lambda t}$，这是顾客到达的间隔时间$T$不超过$t$的概率，即$1 - \mathrm{e}^{-\lambda t} = F(t) = P\{T \leqslant t\}$，恰好是负指数分布函数的表达式。

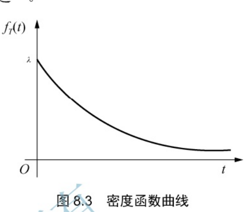


图 8.3　密度函数曲线

泊松过程与负指数分布的关系描述如下。

定理 8.1　当输入过程为泊松过程(或在一定时间内到达系统的顾客数服从泊松分布)时，则顾客相继到达的间隔时间必然服从负指数分布。

间隔时间负指数分布的参数λ也称为平均到达率，即单位时间到达的顾客数，这是因为其数学期望$1/\lambda$是顾客到达的平均间隔时间。当负指数分布用来描述服务时间分布时，参数通常用μ表示，其数学期望$1/\mu$就表示平均服务时间，μ也称为平均服务率，即单位时间服务完的顾客数。

3. *爱尔朗分布*

定义 8.3 爱尔朗分布：若随机变量$T_1, T_2, \cdots, T_k$是相互独立的且服从相同参数k, μ的负指数分布，那么随机变量$T = T_1 + T_2 + \cdots + T_k$服从参数为$k, \mu$的爱尔朗(Erlang)分布，记为$T \sim \mathrm{E}_k(\mu)$，称为$k$阶爱尔朗分布，密度函数

$$f(t) = \frac{k\mu(k\mu t)^{k-1}}{(k-1)!}\mathrm{e}^{-k\mu t} \quad (t \geqslant 0)$$

根据分布的独立性，应用负指数分布的期望和方差，可知爱尔朗分布的数学期望和方差分别为

$$\mathrm{E}(T) = \sum_{i=1}^{k}\mathrm{E}(T_i) = \frac{1}{\mu}; \quad \mathrm{D}(T) = \sum_{i=1}^{k}\mathrm{D}(T_i) = \frac{1}{k\mu^2}$$

爱尔朗分布是随k而变化的一族分布。当$k = 1$时就是负指数分布；当$k \to \infty$时就化为确定性分布；当$1 < k < \infty$时，它是介于随机和确定之间的中间型分布，能对现实世界表现出更好的适应性，如图 8.4 所示。在排队系统中，有些时间就是服从爱尔朗分布，如顾客在系统中要接受k次串列的服务。若每个服务台的服务时间都服从参数为μ的负指数分布，则接受服务的总时间就服从k阶爱尔朗分布。又如在单服务台的排队系统中，当某一个顾客进入系统发现已有n个顾客的条件下，则他的等待时间就服从n阶爱尔朗分布，他在系统的逗留时间就服从$n+1$阶爱尔朗分布。

【例 8.5】港口集装箱物流是指集装箱从到达港口码头至离开港口码头中形成的物流过程。集装箱船到港后，集装箱经过“进入泊位—桥吊卸货—水平运输—跨车/龙门吊卸下—堆场堆存—码头内部运输—其他运输方式转运”流程离开码头。表 8-2 为集装箱装卸排队系统某天的作业数据，若假设到达间隔时间和服务时间都服从负指数分布，试求其平均到达率和服务率。

北京大学出版社版权所有 禁止转载

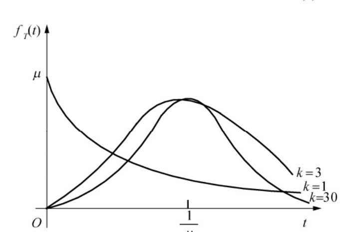


图 8.4　爱尔朗分布密度函数曲线

表 8-2　港口集装箱装卸排队系统

集装箱	到达时间	到达时间间隔/分钟	开始装卸时间	装卸服务时间/分钟	装卸结束时间
1	9：03		9：03	17	9：20
2	9：15	12	9：20	21	8：41
3	9：25	10	8：41	19	10：00
4	9：30	5	10：00	15	10：15
5	10：05	35	10：15	20	10：35
6	—	—	—	—	—

北京大学出版社版权所有 禁止转载

解：根据表中的数据进行到达时间间隔和装卸服务时间统计，计算其平均数就是平均到达率和服务率的估计参数，其理论详见统计学的知识。限于篇幅，表 8-2 的数据没有全部列出，仅就给出的数据进行计算。

平均到达间隔时间 $\frac{1}{\lambda}=\frac{(12+10+5+35)\text{分钟}}{4}=15.5$ 分钟，平均到达率 $\lambda=\frac{1}{15.5}$ 次/分钟；

平均服务时间 $\frac{1}{\mu}=\frac{(17+21+19+15+20)}{5}=18.4$ 分钟，平均服务率 $\mu=\frac{1}{18.4}$ 次/分钟。

在后面的内容中，一般要求平均到达率小于平均服务率，否则排队过程会一直继续下去，不符合排队系统的客观实际。该题出现这个问题是因为列表数据较少，不具有很明确的统计意义，仅用来示例这两个参数的具体求法。

8.1.5　排队系统的绩效指标

在研究一个排队系统时，不仅要分析系统的组成，结构及运行方式，更重要的是衡量一个排队系统的好坏并对其进行经济分析。这就需要一系列描述排队系统特征的数量指标，分析其运行效率，考核服务质量，以便据此提出改进措施，优化服务系统。通常评价排队系统优劣有 6 项数量指标。

(1) 队长。系统中排队等待服务和正在服务的顾客总数，其平均值记为 L_s。

(2) 队列长。系统中排队等待服务的顾客数，其平均值记为 L_q。

(3) 逗留时间。某顾客在系统中的停留时间，包括等待时间和服务时间，其平均值记为 W_s。

(4) 等待时间。某顾客在系统中的排队等待时间，其平均值记为 W_q。

(5) 系统负荷水平 ρ。它是衡量服务台在承担服务和满足需要方面能力的尺度。

(6) 系统状态概率 P_n。系统中有 n 个顾客要求服务的概率，其中 P_0 为系统空闲的概率。

这些测度通常是以期望价值的形式所表达的，是一种系统稳定运行下对系统状态的考量。在排队系统最初的开始阶段，系统中没有顾客，需要经过一段时间系统中的期望顾客数才能达到正常水平，此后这个系统就处于稳定状态中。

8.2 负指数分布排队系统模型

在本节集中讨论输入过程服从泊松过程、服务时间服从负指数分布的排队系统，将其分为单服务台和多服务台排队系统分别讨论。

8.2.1 单服务台排队系统模型

单服务台负指数分布排队系统按下面三种情况分析：①标准的 $M/M/1$ 模型，即 $M/M/1/\infty/\infty$；②系统容量有限制，即 $M/M/1/N/\infty$；③顾客源为有限，即 $M/M/1/\infty/m$。其他类型的排队系统的各种指标计算公式则复杂得多，现已开始应用计算机仿真来求解排队系统问题。

1. 标准的 $M/M/1$ 模型

$M/M/1$ 排队系统是一类最简单、最基本的排队系统。系统的输入过程：顾客到达的间隔时间服从参数为 λ 的负指数分布，顾客源无限。排队规则：单队列，队长无限制，先到先服务。服务机构：单服务台，服务时间服从参数为 μ 的负指数分布，此外还假设服务时间和顾客相继到达的间隔时间相互独立。

1) 系统的状态概率

为明确系统的运行情况，需要了解当前系统中的服务状态，需引入一个随机变量来描述系统中顾客的数量，记系统顾客数量为 n 的概率为 $P_n\ (n=0,1,2,\cdots)$，这样系统在达到稳定情况下的状态转移如图 8.5 所示。

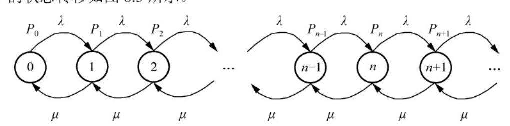


图 8.5 $M/M/1$ 模型系统的状态转移图

图中ⓝ表示状态 $n\ (n=1,2,\cdots)$，从状态 n 变化到状态 $n+1$，是因为顾客的到来，而顾客的到达率为 λ，故在图中表示相邻状态增加的箭线上标上 λ；由状态 $n+1$ 转移到状态 n，是由于顾客的离去，而系统的服务率为 μ，故在图中表示相邻状态减少的箭线上标上 μ。

在系统运行达到稳定后，各种状态进入了动态平衡。首先考虑系统没有顾客这个状态⓪，根据泊松过程的普通性，在某个时间间隔进入系统人数大于 1 的概率是一个无穷小量，可以认定某一时间间隔进入系统的人数至多为 1，若有顾客进入，这时系统中的人数就变为 1，系统状态转化为①。基于泊松分布的平稳性，这样顾客进入系统概率就是 $\lambda\Delta t$，因

北京大学出版社版权所有 禁止转载

为系统现在的状态概率为 P_0，所以系统状态⓪向状态①转移的概率就是 $\lambda\Delta tP_0$；同样顾客离开系统的概率为 $\mu\Delta t$，系统在状态①的概率为 P_1，所以系统状态①向状态⓪转移的概率就为 $\mu\Delta tP_1$，因为动态平衡下进入和流出这个状态的顾客应当持平，就有 $\lambda\Delta tP_0=\mu\Delta tP_1$，即 $\lambda P_0=\mu P_1$。同理，根据状态转移图 8.5，可以得到以下状态转移方程组

$$\begin{cases}\lambda P_0=\mu P_1\\ \lambda P_{n-1}+\mu P_{n+1}=\lambda P_n+\mu P_n \quad n=1,2,\cdots\end{cases} \tag{8-2}$$

求解方程组(8-2)，得

$$P_1=\frac{\lambda}{\mu}P_0$$

$$P_2=-\frac{\lambda}{\mu}P_0+\left(1+\frac{\lambda}{\mu}\right)P_1=\left(\frac{\lambda}{\mu}\right)^2P_0$$

$$P_n=\left(\frac{\lambda}{\mu}\right)^nP_0 \quad (n=1,2,\cdots)$$

设 $\rho=\dfrac{\lambda}{\mu}<1$，则

$$P_1=\rho P_0，\ P_2=\rho^2P_0，\ \cdots，\ P_n=\rho^nP_0$$

由于 $\sum\limits_{n=0}^{\infty}P_n=1$，得

$$\begin{cases}P_0=1-\rho\\ P_n=(1-\rho)\rho^n \quad n=1,2,\cdots\end{cases} \tag{8-3}$$

式(8-3)中：ρ 有明确的实际意义，因为 ρ 是平均到达率和平均服务率之比，且 $\rho=1-P_0$ 表示系统正在服务的概率，因此 ρ 反映了当前排队系统的服务强度或服务机构的利用率。

2) 系统的绩效指标

有了状态概率的解析式，就可以计算排队系统的绩效指标。

L_s 表示系统的平均顾客数，即

$$L_s=\sum_{k=0}^{\infty}kP_k=\sum_{k=0}^{\infty}k\rho^k(1-\rho)=(1-\rho)\sum_{k=0}^{\infty}k\rho^k=(1-\rho)\frac{\rho}{(1-\rho)^2}=\frac{\rho}{1-\rho}$$

为系统中顾客数的期望值。

L_q 表示队列中平均顾客数，即

$$L_q=\sum_{k=1}^{\infty}(k-1)P_k=\sum_{k=1}^{\infty}(k-1)\rho^k(1-\rho)=\rho(1-\rho)\sum_{k=1}^{\infty}(k-1)\rho^{k-1}=\rho(1-\rho)\frac{\rho}{(1-\rho)^2}=\frac{\rho^2}{1-\rho}$$

显然，$L_s-L_q=\rho$。

当顾客到达的间隔时间服从参数为 λ 的负指数分布，顾客在系统中接受服务时间服从参数为 μ 的负指数分布时，由于负指数分布的无记忆性，可以证明顾客在系统中逗留的时间服从参数为 $\mu-\lambda$ 的负指数分布，根据负指数分布的数学期望公式，得

$$W_s=\frac{1}{\mu-\lambda}$$

北京大学出版社版权所有
禁止转载

由于顾客在系统中的逗留时间包含顾客在队列中的等待时间和顾客接受服务时间，所以顾客的等待时间$W_q = W_s - \frac{1}{\mu}$，其中$\frac{1}{\mu}$是顾客服务时间的数学期望，得

$$W_q = W_s - \frac{1}{\mu} = \frac{1}{\mu-\lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{\rho}{\mu-\lambda} = \rho W_s$$

3) 绩效指标之间的关系

从上述4个指标的表达式可得其关系为

$$\begin{aligned} L_s &= \lambda W_s \qquad & L_q &= \lambda W_q \\ L_q &= L_s - \frac{\lambda}{\mu} \qquad & W_q &= W_s - \frac{1}{\mu} \end{aligned} \tag{8-4}$$

这一组公式就是著名的里特(Little)公式，这些关系在其他排队系统中稍加调整也成立。该公式具有非常直观的含义：若系统处于稳定状态，那么系统中的人数等于顾客在系统的平均逗留时间乘以系统的平均到达率。

北京大学出版社版权所有 禁止转载

【例 8.6】 轻轨进站口有一个售票窗口，乘客到达服从泊松分布，平均到达率为200人/小时，售票时间服从负指数分布，平均售票时间为15秒/人。试求：(1)系统空闲的概率；(2)至少有两个顾客排队的概率；(3)该排队系统的有关指标。

解：系统是一个标准的$M/M/1$排队系统，据题知，$\lambda = 200$人/小时，$\mu = \frac{1}{15\text{秒/人}} \times 3\,600\text{秒/小时} = 240$人/小时，则

$$\rho = \frac{\lambda}{\mu} = \frac{200}{240} = \frac{5}{6}。$$

(1) 系统空闲的概率 $P_0 = 1-\rho = 1-\frac{5}{6} = \frac{1}{6}$；

(2) 至少有两个顾客排队的概率为 $1-P_0-P_1 = 1-P_0-\rho P_0 = 1-\frac{1}{6}-\frac{5}{6}\times\frac{1}{6} = \frac{25}{36}$；

(3) 系统中的平均人数 $L_s = \frac{\rho}{1-\rho} = \frac{5/6}{1-5/6} = 5$个；

排队等待的平均人数 $L_q = L_s - \rho = \left(5-\frac{5}{6}\right)$个$\approx 4.17$个；

乘客在系统中的平均逗留时间 $W_s = \frac{1\text{人}}{(240-200)\text{人/小时}} = 0.025\text{小时} = 90\text{秒}$；

乘客在系统中的平均等待时间 $W_q = W_s - \frac{1}{\mu} = 75$秒。

【例 8.7】 考虑一个铁路列车编组站，设待编列车到达时间间隔服从负指数分布，平均每小时到达2列，编组站的编组时间也服从负指数分布，平均每20分钟可编一组。试求该编组系统中列车的平均数，每一辆列车平均停留时间和等待编组的列车数。

解：这是一个标准的$M/M/1$排队系统。

据题知$\lambda = 2$列/小时，$\mu = \frac{1}{20}$列/分=3列/小时，所以$\rho = \frac{\lambda}{\mu} = \frac{2}{3}$。

编组系统中列车的平均数为

$$L_s = \frac{\rho}{1-\rho} = \frac{2}{3}\bigg/\left(1-\frac{2}{3}\right) = 2\,(\text{列});$$

列车的平均停留时间为

$$W_s = \frac{1}{\mu-\lambda} = \frac{1}{3-2} = 1\,(\text{小时});$$

等待列车编组的平均数为

$$L_q = L_s - \rho = 2 - \frac{2}{3} = \frac{4}{3}\,(\text{列})。$$

2. 系统容量有限制的 $M/M/1/N$ 模型

系统容量由无限转化为有限制容量 N 时，模型由 $M/M/1$ 转变为 $M/M/1/N$。若系统已有 N 个顾客，再来的顾客将会被拒绝而不能进入系统。

1) 系统的状态概率

根据服务系统的特征，其状态转移率如图 8.6 所示。

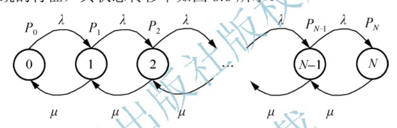


图 8.6 $M/M/1/N$ 模型系统状态转移率图

根据状态转移率图，列出系统状态转移方程组为

$$\begin{cases} \lambda P_0 = \mu P_1 \\ \lambda P_{n-1} + \mu P_{n+1} = (\lambda+\mu)P_n & n = 1, 2, \cdots, N-1 \\ \mu P_N = \lambda P_{N-1} \end{cases}$$

求解方程组，可得 $P_n = \dfrac{\lambda}{\mu}P_{n-1} = \left(\dfrac{\lambda}{\mu}\right)^n P_0$。仍令 $\rho = \lambda/\mu$，当 $\rho \neq 1$ 时，由于 $\sum\limits_{n=0}^{N} P_n = 1$，有

$$\sum_{n=0}^{N} P_n = P_0(1+\rho+\rho^2+\cdots+\rho^N) = P_0\frac{1-\rho^{N+1}}{1-\rho}$$

解得

$$\begin{cases} P_0 = \dfrac{1-\rho}{1-\rho^{N+1}} & \rho \neq 1 \\ P_n = \rho^n P_0 & n = 1, 2, \cdots, N \end{cases}$$

当 $\rho = 1$ 时，$P_N = P_{N-1} = \cdots = P_0 = \dfrac{1}{N+1}$。

由于系统容量有限制，队伍不会排无限长。状态概率 P_N 表示系统的满员率，从离开的角度来考虑，也可以称为系统对顾客的拒绝率或顾客的损失率。

2) 系统的绩效指标

先来求系统中的平均队长，当 $\rho \neq 1$ 时，得

$$L_s = \sum_{n=0}^{N} nP_n = P_0\sum_{n=0}^{N} n\rho^n = \frac{\rho}{1-\rho} - \frac{(N+1)\rho^{N+1}}{1-\rho^{N+1}}$$

北京大学出版社版权所有 禁止转载

当 $\rho=1$ 时，$L_s=\sum_{n=0}^{N} nP_n=\frac{1}{N+1}(1+2+\cdots+N)=\frac{N}{2}$

平均队列长 $L_q=\sum_{k=1}^{N}(k-1)P_k=\sum_{k=1}^{N}kP_k-\sum_{k=1}^{N}P_k=L_s-(1-P_0)$

模型 $M/M/1/N$ 与标准模型 $M/M/1$ 的区别有两个：一是系统的状态只有 $N+1$ 个；二是顾客实际进入系统率，当 $n<N$ 时为 λ，当 $n=N$ 时为 0，故平均值(有效到达率) $\lambda_e=\lambda(1-P_N)+0P_N=\lambda(1-P_N)$，即单位时间内到达并能进入队列的平均数，从服务的角度来考虑有效到达率为 $\mu(1-P_0)$，即表示系统不空且以速率 μ 提供服务。

若令 $\rho_e=\frac{\lambda_e}{\mu}$，$\rho_e$ 称为有效服务强度，这样就有 $L_q=L_s-\rho_e$。

平均逗留时间 $W_s=\frac{L_s}{\lambda_e}=\frac{L_s}{\lambda(1-P_N)}$。

平均等待时间 $W_q=W_s-\frac{1}{\mu}$。

$M/M/1/N$ 排队系统模型的绩效指标归纳为

$$
\begin{aligned}
&L_s=\begin{cases}\dfrac{\rho}{1-\rho}-\dfrac{(N+1)\rho^{N+1}}{1-\rho^{N+1}}, & \rho\neq 1\\[2mm] \dfrac{N}{2}, & \rho=1\end{cases}\\
&L_q=L_s-\rho_e,\ \lambda_e=\lambda(1-P_N)=\mu(1-P_0)\\
&W_s=\frac{L_s}{\lambda_e}=\frac{L_s}{\lambda(1-P_N)}\\
&W_q=W_s-\frac{1}{\mu}
\end{aligned}
\tag{8-5}
$$

【例 8.8】 某火车货运站以平均每天 4 列的泊松流到达车站，车站的装卸能力是每天 8 列。站内除了装卸站台外，另有两股道可供停车等待。求

(1) 货车一到站就能装卸的概率和不能进站的概率；

(2) 货运站的有效到达率；

(3) 站内的平均列车数和平均等待的列车数；

(4) 货车在站内的平均逗留时间和平均等待时间；

(5) 装卸站台的利用率。

解：这是一个系统容量有限制的模型，$N=3$，$\lambda=4$ 列/天，$\mu=8$ 列/天，则 $\rho=\frac{\lambda}{\mu}=\frac{1}{2}=0.5$。

(1) 货车一到站就能装卸的概率 $P_0=\frac{1-0.5}{1-0.5^4}=0.53$，货车不能进站的概率为

$$P_3=\rho^3\ P_0=0.5^3\times 0.53\approx 0.07$$

(2) 货运站的有效到达率 $\lambda_e=\lambda(1-P_3)=4\times(1-0.07)$列/天$=3.72$ 列/天；

(3) $L_s=\frac{0.5}{1-0.5}-\frac{4\times 0.5^4}{1-0.5^4}=0.73$ (列)，$L_q=[0.73-(1-0.53)]$列$=0.26$ 列；

北京大学出版社版权所有 禁止转载

(4) $W_s=\dfrac{0.73\text{列}}{3.72\text{列/天}}=0.20$ 天，　$W_q=W_s-\dfrac{1}{\mu}=\left(0.20-\dfrac{1}{8}\right)$天 ≈ 0.08 天；

(5) 装卸站台的利用率　$1-P_0=1-0.53=0.47$。

3. 顾客源有限的 $M/M/1/\infty/m$ 模型

有限顾客源模型中顾客的总数为 m，当顾客需要服务时，就进入队列等待；服务完毕后，重新回到顾客源中，如此循环往复，如图 8.7 所示。顾客源有限的排队系统最典型的问题是机器维修问题，机器出现故障表示顾客“到达”，待维修的机器形成队列，修理工人是服务台。模型中系统容量是“∞”，实际系统的容量永远也不超过 m，系统的符号实际可表示为 $M/M/1/m/m$。

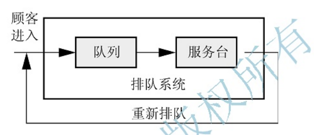


图 8.7　有限顾客源排队系统

北京大学出版社版权所有 禁止转载

在无限顾客源系统中，顾客的到达速率 λ 是整个顾客源的特点，与单独的顾客无关。而在有限顾客源系统中，由于一个顾客要反复接受服务，因此有必要假定每一个顾客在单位时间内需要接受服务的平均次数是相同的，仍设为 λ。这样系统中顾客到达的平均速率就与顾客源的顾客数有关，仍以机器维修为例，设机器总数为 m 台，每台机器在单位时间内发生故障的平均次数为 λ，已经发生故障正在等待维修及正在接受维修的机器数为 n，则在单位时间内出现故障的平均机器数(即有限顾客源的平均到达率)为 $\lambda(m-n)$。

1) 系统的状态概率

根据服务系统的特征，其状态转移率如图 8.8 所示。

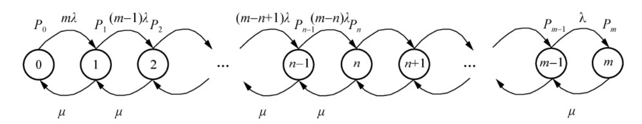


图 8.8　$M/M/1/m/m$ 模型系统状态转移率图

根据状态转移率图，列出系统状态转移方程组为

$$\begin{cases} m\lambda P_0=\mu P_1 \\ (m-n+1)\lambda P_{n-1}+\mu P_{n+1}=[(m-n)\lambda+\mu]P_n \qquad n=1,2,\cdots,m-1 \\ \mu P_m=\lambda P_{m-1} \end{cases}$$

仍令 $\rho=\dfrac{\lambda}{\mu}$，由于 $\sum\limits_{n=0}^{m}P_n=1$，求解方程组，得

$$\begin{cases} P_0 = \dfrac{1}{\sum\limits_{i=0}^{m} \dfrac{m!}{(m-i)!}\rho^i} \\ P_n = \dfrac{m!}{(m-n)!}\rho^n P_0 \qquad n = 1, 2, \cdots, m \end{cases}$$

2) 系统的绩效指标

由 Little 公式，得

$$L_s - L_q = \frac{\lambda_e}{\mu} = \frac{\lambda(m - L_s)}{\mu}$$

另一方面

$$L_s - L_q = 1 - P_0$$

结合上面两式，解得

$$L_s = m - \frac{\mu}{\lambda}(1 - P_0)$$

$\lambda(m - L_s) = \mu(1 - P_0)$ 的意义是很直观的：$m - L_s$ 表示系统外部的顾客数，$\lambda_e = \lambda(m - L_s)$ 表示系统的有效到达率；$1 - P_0$ 是系统忙的概率，$\mu(1 - P_0)$ 为系统忙时的服务率，故二者应当相等。

最后得到系统的绩效指标为

$$\begin{cases} L_s = m - \dfrac{\mu}{\lambda}(1 - P_0) \\ L_q = L_s - (1 - P_0) \\ W_s = \dfrac{L_s}{\mu(1 - P_0)} = \dfrac{m}{\mu(1 - P_0)} - \dfrac{1}{\lambda} \\ W_q = W_s - \dfrac{1}{\mu} \end{cases} \tag{8-6}$$

【例 8.9】 有 5 辆叉车往返于仓库之中理货，每辆叉车的连续运行时间服从负指数分布，每天(8 小时)平均连续运行时间 300 分钟，配备 1 个修理工，修理时间也服从负指数分布，平均修理时间为 240 分钟。试求：

(1) 修理工空闲的概率和 5 辆叉车都出故障的概率；

(2) 叉车的平均故障辆数；

(3) 叉车等待修理的平均时间；

(4) 修理工的平均利用率。

解：这是一个有限顾客源的模型，$m = 5$，$\lambda = 1.6$ 辆/天，$\mu = 2$ 辆/天，则 $\rho = \dfrac{\lambda}{\mu} = \dfrac{1.6}{2} = 0.8$，根据公式(8-6)，计算得

(1) 修理工空闲的概率为

$$P_0 = \left[\frac{5!}{5!}(0.8)^0 + \frac{5!}{4!}(0.8)^1 + \frac{5!}{3!}(0.8)^2 + \frac{5!}{2!}(0.8)^3 + \frac{5!}{1!}(0.8)^4 + \frac{5!}{0!}(0.8)^5\right]^{-1} = 0.007\,3$$

5 辆叉车都出故障的概率 $\quad P_5 = \dfrac{5!}{0!}(0.8)^5 P_0 = 0.287$；

北京大学出版社版权所有 禁止转载

(2) 叉车的平均故障辆数

$$L_s = m - \frac{\mu}{\lambda}(1-P_0) = 5 - \frac{1}{0.8}(1-0.007\,3) = 3.76\ (\text{台});$$

(3) 叉车等待修理的平均时间为

$$W_q = \frac{L_s}{\mu(1-P_0)} - \frac{1}{\mu} = \frac{3.76}{2\times(1-0.007\,3)} - \frac{1}{2} = 1.39\ (\text{天});$$

(4) 修理工的平均利用率 $1-P_0 = 0.9927$。

基于上述数据，说明系统的利用率太高，维修工人劳动强度太大，没有休息时间，建议增加维修人员或提高维修工人待遇。

8.2.2 多服务台排队系统模型

多服务台负指数分布排队系统分为：①标准的 $M/M/c$ 模型；②系统容量有限制，即 $M/M/c/N/\infty$；③顾客源为有限，即 $M/M/c/\infty/m$。

1. 标准的 $M/M/c$ 模型

标准的多服务台模型与标准的单服务台模型的条件相同，除了服务台数量不同。为便于讨论，设各服务台工作独立且服务速率相同，于是整个系统的平均服务速率为 $c\mu$。

1) 系统的状态概率

这个系统的特点为系统的服务速率与系统中的顾客数无关。当系统中的顾客数 k 不大于服务台个数时，系统中的顾客全部在服务台中，这时系统的服务速率是 $k\mu$；当系统中顾客数大于服务台数时，服务台中正在接受服务的顾客数仍为 c 个，其余顾客在队列中等待服务，这时系统的服务速率为 $c\mu$，其状态转移率如图 8.9 所示。

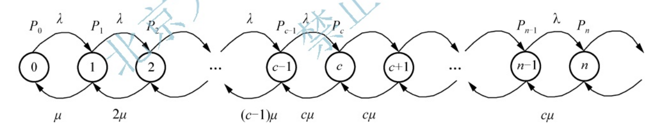


图 8.9 $M/M/c$ 模型系统状态转移率

根据状态转移率图，列出系统状态转移方程组为

$$\begin{cases} \lambda P_0 = \mu P_1 \\ \lambda P_{n-1} + (n+1)\mu P_{n+1} = (\lambda + n\mu)P_n & n < c \\ \lambda P_{n-1} + c\mu P_{n+1} = (\lambda + c\mu)P_n & n \geqslant c \end{cases}$$

令 $\rho = \dfrac{\lambda}{c\mu} < 1$，由于 $\sum\limits_{n=0}^{\infty} P_n = 1$，求解方程组，得

$$P_0 = \left[\sum_{k=0}^{c-1}\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k + \frac{1}{c!(1-\rho)}\left(\frac{\lambda}{\mu}\right)^c\right]^{-1}$$

北京大学出版社版权所有 禁止截获

$$P_n = \begin{cases} \dfrac{1}{n!}\left(\dfrac{\lambda}{\mu}\right)^n P_0 & n \leqslant c \\ \dfrac{1}{c!c^{n-c}}\left(\dfrac{\lambda}{\mu}\right)^n P_0 & n > c \end{cases}$$

2) 系统的绩效指标

系统的绩效指标为

$$\begin{cases} L_s = L_q + \dfrac{\lambda}{\mu} \\ L_q = \dfrac{(c\rho)^c \rho}{c!(1-\rho)^2} P_0 \qquad \rho = \dfrac{\lambda}{c\mu} < 1 \\ W_s = \dfrac{L_s}{\lambda} \\ W_q = \dfrac{L_q}{\lambda} = W_s - \dfrac{1}{\mu} \end{cases} \tag{8-7}$$

【例 8.10】 某售票站有 3 个售票窗口，顾客到达服从泊松流，均值为 0.9 人/分钟，售票时间服从负指数分布，平均服务率 0.4 人/分钟。如果排队方式是顾客到达后排出一队，依次向空闲的窗口购票，试求：

(1) 售票站空闲的概率；

(2) 售票站内的平均顾客数和排队等待的顾客数；

(3) 顾客购票的平均消耗时间和平均排队等待的时间；

(4) 顾客到达后必须排队的概率。

解： 这是一个标准的 $M/M/3$ 排队系统，$\lambda = 0.9$ 人/分钟，$\mu = 0.4$ 人/分钟，$\dfrac{\lambda}{\mu} = \dfrac{0.9}{0.4} = 2.25$，$\rho = \dfrac{\lambda}{c\mu} = \dfrac{0.9}{3 \times 0.4} = 0.75$。

(1) 售票站空闲的概率为

$$P_0 = \left[\sum_{k=0}^{c-1} \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k + \frac{1}{(c)!(1-\rho)}\left(\frac{\lambda}{\mu}\right)^c\right]^{-1}$$

$$= \left[\frac{(2.25)^0}{0!} + \frac{(2.25)^1}{1!} + \frac{(2.25)^2}{2!} + \frac{(2.25)^3}{3! \times (1-0.75)}\right]^{-1} = 0.074\,8\text{；}$$

(2) 售票站内排队等待的顾客数为

$$L_q = \frac{(c\rho)^c \rho}{c!(1-\rho)^2} P_0 = \frac{(2.25)^3 \times 0.75}{3! \times (1-0.75)^2} \times 0.074\,8 = 1.70\,(\text{人})\text{；}$$

售票站内的平均顾客数为

$$L_s = L_q + \frac{\mu}{\lambda} = 1,\ (70 + 2.25)\text{人} = 3.95\text{ 人；}$$

(3) 顾客的平均消耗时间为

$$W_s = \frac{L_s}{\lambda} = \frac{3.95\text{人}}{0.9\text{人/分钟}} = 4.39\text{ 分钟；}$$

北京大学出版社版权所有 禁止转载

顾客的平均等待时间为

$$W_q = \frac{L_q}{\lambda} = \frac{1.70\text{人}}{0.9\text{人/分钟}} = 1.89\text{分钟};$$

(4) 顾客到达后必须排队的概率为

$$P(n \geqslant 3) = 1 - P_0 - P_1 - P_2 = 1 - 0.074\,8 - 2.25 \times 0.074\,8 - \frac{1}{2} \times (2.25)^2 \times 0.074\,8 = 0.57$$

在例 8.10 中，如果顾客的排队方式变为到达售票站后可到任意窗口排队，且入队后不再换队，即可形成 3 个队列。这时，原来的 $M/M/3$ 模型就变成了 3 个 $M/M/1$ 模型，且每个子系统的平均到达率为 $\lambda_1 = \lambda_2 = \lambda_3 = 0.3$ (人/分钟)，按 $M/M/1$ 模型重新求解本例的问题，并将结果列于表 8-3 进行比较。

表 8-3　$M/M/3$ 模型与 $M/M/1$ 模型指标对比

指标＼模型	$M/M/3$	3 个 $M/M/1$
服务台空闲的概率	0.074 8	0.25(每个)
顾客必须等待的概率	0.57	0.75
平均队长 L_s/人	3.95(整个系统)	9(每个窗口)
平均队列长 L_q/人	1.70	2.25(每个窗口前)
平均逗留时间 W_s/分钟	4.39	10
平均排队时间 W_q/分钟	1.89	7.5

从表 8-3 中可知，各项指标都是 $M/M/3$ 模型比 3 个 $M/M/1$ 模型有显著的优越性，服务质量更好，尤其是顾客平均等待时间。因此科学地运用排队规则，在不增加设施成本的情况下，也会大大提高服务质量。

2. 系统容量有限制 $M/M/c/N$ 模型

设系统最大容量为 $N(N > c)$，当系统内的顾客数 $n \leqslant N$ 时，到达的顾客进入系统；当 $n > N$ 时，到达的顾客被拒绝，如图 8.10 所示。设顾客的到达率为 λ，每个服务台的服务率为 μ，$\rho = \lambda / c\mu$。由于系统不会无限地接纳顾客，对 ρ 不必加以限制。

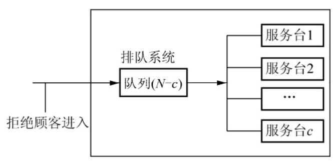


图 8.10　系统容量有限排队系统

1) 系统的状态概率

系统的状态转移率如图 8.11 所示，这时状态个数为有限值。

北京大学出版社版权所有 禁止转载

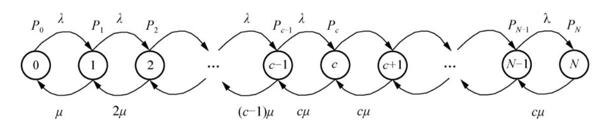


图 8.11 *M*/*M*/*c*/*N* 模型系统状态转移率

根据状态转移率图，列出系统状态转移方程组为

$$\begin{cases}\lambda P_0 = \mu P_1 \\ \lambda P_{n-1} + (n+1)\mu P_{n+1} = (\lambda + n\mu)P_n & n < c \\ \lambda P_{n-1} + c\mu P_{n+1} = (\lambda + c\mu)P_n & c \leqslant n < N \\ \lambda P_{N-1} = c\mu P_N \end{cases}$$

求解方程组，得

$$P_0 = \left[\sum_{k=0}^{c}\frac{1}{k!}(c\rho)^k + \frac{c^c}{c!}\frac{\rho(\rho^c - \rho^N)}{1-\rho}\right]^{-1}$$

$$P_n = \begin{cases}\dfrac{1}{n!}(c\rho)^n P_0 & n \leqslant c \\ \dfrac{c^c}{c!}(\rho)^n P_0 & c \leqslant n \leqslant N \end{cases}$$

2) 系统的绩效指标

系统的绩效指标为

$$\begin{cases} L_q = \dfrac{\rho(c\rho)^c}{c!(c-\rho)^2}\left[1-\rho^{N-c} - (N-c)\rho^{N-c}(1-\rho)\right]P_0 \\ L_s = L_q + c\rho(1-P_N) \\ W_s = \dfrac{L_s}{\lambda(1-P_N)} \\ W_q = W_s - \dfrac{1}{\mu} \end{cases} \tag{8-8}$$

特别当 $N=c$ 时为即时制(损失制)的情形，如街头停车场、旅店床位等，系统容量与服务台相等，如果服务台没有空，顾客当即离去，这样的系统成为“即时制”。

【例 8.11】某旅馆有 10 个床位，顾客到达服从泊松流，平均速率为 6 人/天，旅客平均逗留时间为 2 天，试求：(1)旅馆客满的概率；(2)每天客房平均占用数。

解：根据题知，这是一个即时制的多服务台服务系统，其中

$$N = c = 10,\quad \lambda = 6,\quad \mu = 0.5,\quad \frac{1}{\mu} = 2,\quad c\rho = \frac{6}{0.5} = 12$$

(1) 旅馆 10 个床位全满的概率为

$$P_0 = \left[\sum_{k=0}^{c}\frac{1}{k!}(c\rho)^k + \frac{c^c}{(c)!}\frac{\rho(\rho^s - \rho^N)}{1-\rho}\right]^{-1}$$

北京大学出版社版权所有 禁止转载

$$=\left[\frac{1}{0!}(12)^0+\frac{1}{1!}(12)^1+\cdots+\frac{1}{10!}(12)^{10}+\frac{c^c}{(c)!}\frac{\rho(\rho^c-\rho^c)}{1-\rho}\right]^{-1}$$
$$=0.0018$$

$$P_{10}=\frac{1}{10!}(c\rho)^n P_0=\frac{1}{10!}(12)^{10}\times 0.0018=0.3019$$

(2) 每天客房平均占用数为

$$L_s=L_q+c\rho(1-P_N)=12\times(1-0.3019)=8.3772\,(\text{床位})$$

即平均占用 8.377 个床位，客房占用率为 83.77%。

3. 顾客源有限 $M/M/c/\infty/m$ 模型

有限顾客源模型中顾客的总数为 m，服务台个数为 c，且 $m>c$。顾客源有限的排队系统最典型的问题是机器维修问题，机器数量 m 台，修理工数量为 c 人。

1) 系统的状态概率

根据服务系统的特征，其状态转移率如图 8.12 所示。

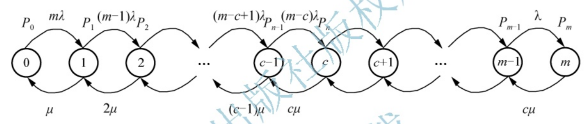


图 8.12　$M/M/c/\infty/m$ 模型系统状态转移率

根据状态转移率图，列出系统状态转移方程组，求解方程组，得

$$P_0=\frac{1}{m!}\left[\sum_{k=0}^{c}\frac{1}{k!(m-k)!}\left(\frac{c\rho}{m}\right)^k+\frac{c^c}{c!}\sum_{k=c+1}^{m}\frac{1}{(m-k)!}\left(\frac{\rho}{m}\right)^k\right]^{-1}$$

$$P_n=\begin{cases}\dfrac{m!}{(m-n)!n!}\left(\dfrac{\lambda}{\mu}\right)^n P_0 & 0\leqslant n\leqslant c\\[2ex] \dfrac{m!}{(m-n)!c!c^{n-c}}\left(\dfrac{\lambda}{\mu}\right)^n P_0 & c+1\leqslant n\leqslant m\end{cases}$$

其中 $\rho=\dfrac{m\lambda}{c\mu}$。

2) 系统的绩效指标

系统的绩效指标为

$$\begin{cases}L_s=L_q+\dfrac{\lambda}{\mu}(m-L_s)\\[1ex] L_q=\displaystyle\sum_{n=c+1}^{m}(n-c)P_n\\[1ex] W_s=\dfrac{L_s}{\lambda_e}=\dfrac{L_s}{\lambda(m-L_s)}\\[1ex] W_q=W_s-\dfrac{1}{\mu}\end{cases}\tag{8-9}$$

北京大学出版社版权所有 禁止转载

【例 8.12】设有两名维修工人，负责 5 台机器的故障维修。每台机器的平均损坏率为每小时 1 次，两工人能以相同的平均修复率 4 次/小时修好机器，试求系统空闲的概率。

解：这是一个有限顾客源的排队模型，$m=5$，$\lambda=1$(台/小时)，μ=4(次/小时)，$c=2$，$c\rho/m=1/4$，则

$$P_0=\frac{1}{m!}\left[\sum_{k=0}^{c}\frac{1}{k!(m-k)!}\left(\frac{c\rho}{m}\right)^k+\frac{c^c}{c!}\sum_{k=c+1}^{m}\frac{1}{(m-k)!}\left(\frac{\rho}{m}\right)^k\right]^{-1}$$

$$=\frac{1}{5!}\left[\frac{1}{5!}\left(\frac{1}{4}\right)^0+\frac{1}{4!}\left(\frac{1}{4}\right)^1+\frac{1}{2!3!}\left(\frac{1}{4}\right)^2+\frac{2^2}{2!2!}\left(\frac{1}{8}\right)^3+\left(\frac{1}{8}\right)^4+\left(\frac{1}{8}\right)^5\right]=0.314\,9$$

8.3 其他服务时间分布排队系统模型

北京大学出版社版权所有 禁止转载

前面探讨了服务时间服从负指数分布的排队系统模型，该节分析服务时间服从其他分布的排队系统特征。

8.3.1 一般分布模型 *M*/*G*/1

M/*G*/1 模型的基本条件是：①输入过程：顾客源是无限的，到达过程是参数为 λ 的泊松过程；②排队规则，单队列，队列长无限制，先到先服务；③服务机构，单服务台，G 表示服务时间 T 的分布，但已知期望 $\mathrm{E}(T)$ 和方差 $\mathrm{D}(T)$。

与 *M*/*M*/1 排队模型相比，*M*/*G*/1 排队模型的区别就是服务时间不再服从负指数分布。令 $\rho=\lambda\mathrm{E}(T)$，当 $\rho<1$ 时，可以证明

$$L_s=\rho+\frac{\rho^2+\lambda^2\mathrm{D}(T)}{2(1-\rho)}$$

该公式又称 P-K 公式(Pollaczek-Khintchine)。只要给出有关参数，不论时间服从什么分布，总可用 P-K 公式求出平均队长，其他绩效指标为

$$\begin{cases}L_s=\rho+\dfrac{\rho^2+\lambda^2\mathrm{D}(T)}{2(1-\rho)}\\ L_q=L_s-\rho\\ W_s=\dfrac{L_s}{\lambda}\\ W_q=W_s-\mathrm{E}(T)\end{cases}\tag{8-10}$$

在公式(8-10)中，若令 $\mathrm{E}(T)=\dfrac{1}{\mu}$，$\mathrm{D}(T)=\dfrac{1}{\mu^2}$(负指数分布时)，就得到 *M*/*M*/1 模型的绩效指标计算公式。

【例 8.13】某维修站有一技工修理故障机器。已知机器按泊松流发生故障，平均故障率为每小时 5 台，机器排队有两种类型：一种是修理时间 9 分钟；一种是 12 分钟，由统计资料知，$2/3$ 的故障需要修理 9 分钟，$1/3$ 的故障需要修理 12 分钟，试求此维修站的绩效指标。

解：服务时间 T 可以看成是二点分布，则

$$\mathrm{E}(T)=\frac{1}{\mu}=\left(9\times\frac{2}{3}+12\times\frac{1}{3}\right)\text{分钟}=10\text{ 分钟}，\quad \mu=6\text{ 台/小时}$$

$$\mathrm{D}(T)=\left(9^2\times\frac{2}{3}+12^2\times\frac{1}{3}-10^2\right)\times\frac{1}{60^2}=\frac{1}{1800}\ (\text{小时}^2)$$

由 $\rho=\frac{5}{6}$，$\lambda=5$，$\mathrm{D}(T)=\frac{1}{1800}$，应用 P-K 公式，得

$$L_s=\rho+\frac{\rho^2+\lambda^2\mathrm{D}(T)}{2(1-\rho)}=\frac{5}{6}+\frac{\left(\frac{5}{6}\right)^2+5^2\times\frac{1}{1800}}{2\times\left(1-\frac{5}{6}\right)}=2.96\ (\text{台})；$$

$$L_q=\frac{\rho^2+\lambda^2\mathrm{D}(T)}{2(1-\rho)}=\frac{\left(\frac{5}{6}\right)^2+5^2\times\frac{1}{1800}}{2\times\left(1-\frac{5}{6}\right)}=2.13\ (\text{台})；$$

$$W_s=\frac{L_s}{\lambda}=\frac{2.96\text{台}}{5\text{台/小时}}=0.59\text{ 小时}；$$

$$W_q=W_s-\frac{1}{\mu}=0.59-\frac{1}{6}=0.43\ (\text{小时})。$$

8.3.2 定长分布模型

该模型表示为 $M/D/1$，D 表示服务时间为固定时间，即为常数，是一般分布时间的一个特例。由于服务时间 T 是常数，所以方差 $\mathrm{D}(T)=0$，代入 P-K 公式就得各个定长排队模型的有关指标。

北京大学出版社版权所有 禁止转载

【例 8.14】某机场有一条供飞机起降的跑道，飞机请求起降的间隔时间服从均值为 10 分钟的负指数分布，若每架飞机起降占用跑道的时间是定长时间 6 分钟，求系统的各项绩效指标。

解：根据题知 $\lambda=\frac{60\text{分钟/小时}}{10\text{分钟/架}}\times 60=6$ 架/小时，$\mathrm{E}(T)=\frac{6\text{分钟}}{60\text{分钟/小时}}=\frac{1}{10}$ 小时，$\mathrm{D}(T)=0$，

$\rho=\lambda\mathrm{E}(T)=6\times\frac{1}{10}=0.6$。应用 P-K 公式，得

$$L_s=\rho+\frac{\rho^2+\lambda^2\mathrm{D}(T)}{2(1-\rho)}=0.6+\frac{(0.6)^2}{2\times(1-0.6)}=1.05\ (\text{架})；$$

$$L_q=L_s-\rho=(1.05-0.6)\text{架}=0.45\text{ 架}；$$

$$W_s=\frac{L_s}{\lambda}=\frac{1.05\text{架}}{6\text{架/小时}}=0.175\text{ 小时}；$$

$$W_q=\frac{L_q}{\lambda}=\frac{0.45\text{架}}{6\text{架/小时}}=0.075\text{ 小时}。$$

8.3.3 爱尔朗分布模型

该模型表示为 $M/E_k/1$，其中每一个顾客必须依次经过 k 个服务台，接受 k 次服务后才构成一个完整的服务过程。该模型假设每个服务台的服务时间 T_i 服从相同的负指数分布，

则总的服务时间为

$$T = \sum_{i=1}^{k} T_i$$

服从k阶爱尔朗分布，其他条件与标准$M/M/1$相同，实际上k个$M/M/1$串联的排队模型，由于各个子系统的独立性，整个串联系统的绩效等于各个子系统绩效指标的和。

令$\mathrm{E}(T_i) = \frac{1}{k\mu}$，$\mathrm{D}(T_i) = \frac{1}{(k\mu)^2}$，则$\mathrm{E}(T) = \frac{1}{\mu}$，$\mathrm{D}(T) = \frac{1}{k\mu^2}$，$\rho = \frac{\lambda}{\mu}$，同前面的讨论，可得模型的绩效指标为

$$\begin{cases} L_s = \rho + \dfrac{\rho^2 + \lambda^2 \dfrac{1}{k\mu^2}}{2(1-\rho)} = \rho + \dfrac{(k+1)\rho^2}{2k(1-\rho)} \\ L_q = L_s - \rho \\ W_s = \dfrac{L_s}{\lambda} \\ W_q = \dfrac{L_q}{\lambda} \end{cases} \tag{8-11}$$

【例 8.15】一个质量检查员平均每小时收到 2 件送来检验的样品，每件样品要依次完成 5 项检验才能判断是否合格。据统计资料，每项检验所需时间的期望值都是 4 分钟，每项检验的时间和送检产品的到达间隔都服从负指数分布，求其检验过程的各项指标。

解：该检验是一个$M/E_5/1$的排队系统，其中

$$\lambda = \frac{2件}{60分钟} = \frac{1}{30}件/分钟 \quad \frac{1}{\mu} = 4分钟/项 \times 5项/件 = 20分钟/件 \quad \rho = \frac{\lambda}{\mu} = \frac{1/30}{1/20} = \frac{20}{30} = \frac{2}{3}$$

则

$$L_s = \frac{2}{3} + \frac{(5+1)\times(2/3)^2}{2\times 5\times(1-2/3)} = \frac{22}{15}(件)；$$

$$L_q = L_s - \rho = \left(\frac{22}{15} - \frac{2}{3}\right)件 = \frac{4}{5}件；$$

$$W_s = \frac{L_s}{\lambda} = \frac{{}^{22}\!/_{15}件}{{}^{1}\!/_{30}件/分钟} = 44分钟；$$

$$W_q = \frac{L_q}{\lambda} = \frac{{}^{4}\!/_{5}件}{{}^{1}\!/_{30}件/分钟} = 24分钟。$$

8.4 排队系统的优化

面对排队拥挤现象，人们总是希望尽量设法减少排队。通常的做法是增加服务设施，但是增加的数量越多，人力、物力的支出就越大，甚至会出现空闲浪费；如果服务设施太少，顾客排队等待的时间就会很长，这样会给顾客带来不良影响。于是，顾客排队时间的长短与服务设施规模的大小，就构成了设计随机服务系统中的一对矛盾，如何做到既保证一定的服务质量指标，又使服务设施费用经济合理，恰当地解决顾客排队时间与

北京大学出版社版权所有 禁止转载

服务设施费用大小这对矛盾，这就是随机服务系统理论——排队系统优化所要研究解决的问题。

评价一个排队系统的好坏要以顾客与服务机构两方面的利益为标准。就顾客来说总希望等待时间或逗留时间越短越好，从而希望服务台个数尽可能多些。但是，就服务机构来说，增加服务台数，就意味着增加投资，增加多了会造成浪费，增加少了要引起顾客的抱怨甚至失去顾客，那到底增加多少比较好呢？顾客与服务机构为了照顾自己的利益对排队系统中的 3 个指标：队长、等待时间、服务台的忙期(简称忙期)都很关心，因此这 3 个指标也就成了排队论的主要研究内容。

8.4.1 排队系统经济分析

在排队问题中，整个系统的经济效益由顾客的等待费用和服务机构的服务费用组成。因问题的不同，顾客的等待费用有的较低，甚至忽略不计，如到商店购买商品的顾客、购买车票的旅客等，这里的等待费用较低是指直接能衡量的损失较低，但隐含着无形的损失是无法用数字来度量的，比如信誉，不愿长时间等待而离去失去销售机会等损失。有的顾客的等待费用是很高的，如等待修理的设备长时间得不到修理，将严重影响生产，这将直接影响企业的经济效益。又如长时间等不到装卸的轮船和火车，其等待损失就更大了。

服务机构的服务费用包括服务台配置的设施费用、能耗费用、服务人员的工资等。显然服务水平越高，要求的服务设施就越先进，服务人员的技术水平越高，或者增加服务台，当然也要花费更多的资金和付更高的工资。另外对于系统容量限制问题，有的系统容量空间的提供无须花费多少钱，如商店、理发馆、售票处等。但有的系统要增加一个容量空间都会花费较多的资金，如火车站供等待装卸的站台、铁路股道，又如海轮码头供轮船等待装卸的泊位等，每增加一个等待空间都会花巨资来建设、维护和管理。

提高服务质量和提高服务机构的效率是一对矛盾，提高服务机构水平(质量、数量)自然会降低顾客的等待时间(顾客的损失)，这是顾客所期望的，但却增加了服务机构的服务成本，最优化的目标之一是使二者费用之和为最小，并确定达到最优的服务水平，如图 8.13 所示，另一个常用的目标函数是使纯收入或利润(服务收入与服务成本之差)最高。排队系统优化问题包括：①确定最优服务率 μ；②确定最佳服务台数量；③选择最为合适的服务规则。

北京大学出版社版权所有 禁止传播

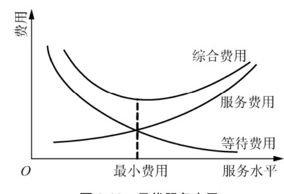


图 8.13 最优服务水平

8.4.2 *M*/*M*/1 系统最优服务率

设 c_s 为系统对每位顾客的单位时间服务成本，c_w 为每位顾客在系统逗留单位时间的损失费用，z 表示总费用。z 由来自系统的服务成本 $c_s\mu$ 和来自顾客的等待损失 $c_w L_s$ 构成，目

标函数是极小化总费用 z，即

$$\min z = c_s\mu + c_w L_s$$

由于 $L_s = \dfrac{\lambda}{\mu-\lambda}$，对上面目标函数求导数，得

$$\frac{\mathrm{d}z}{\mathrm{d}\mu} = c_s - c_w\lambda\frac{1}{(\mu-\lambda)^2} = 0$$

解得

$$\mu^* = \lambda + \sqrt{\frac{c_w}{c_s}\lambda}$$

如果目标函数为系统的服务利润最大化，仍记 z 表示总费用，则表达式为

$$\max z = \mu(1-P_0)G - c_s\mu$$

其中 G 为单位时间对每位顾客服务的收入，P_0 是系统空闲的概率。

【例 8.16】设货船按泊松流到达天津港，平均到达率 $\lambda = 50$ 艘/天，平均卸货率为 μ。已知船在港口停泊一天的费用为一个货币单位，平均卸货费为 μc_s，其中 c_s 是两个货币单位，求使得总费用最少的平均服务率。

解：该问题属于标准的 $M/M/1$ 系统服务率优化问题，其中

$$\lambda = 50, \quad c_s = 2, \quad c_w = 1$$

最佳服务率

$$\mu^* = \lambda + \sqrt{\frac{c_w}{c_s}\lambda} = \left(50 + \sqrt{\frac{1}{2}\times 50}\right)\text{艘/天} = 55\text{艘/天。}$$

8.4.3 *M*/*M*/*c* 系统最优服务设施数

这里仅讨论标准的 $M/M/c$ 系统。设 c_s 为每个服务台的单位时间服务成本，c_w 为每位顾客在系统逗留单位时间的损失费用，z 表示总费用。z 由来自系统的服务成本 $c_s c$ 和来自顾客的等待损失 $c_w L_s$ 构成。目标函数是极小化总费用 z，即

$$\min z = c_s c + c_w L_s$$

这里 $z = z(c)$ 是服务台数量的离散函数，不能采用求导方法求极值点，可采用边际分析法确定服务台的数量。设 c^* 为目标函数的极小点，则 c^* 应满足

$$\begin{cases} z(c^*) \leqslant z(c^*-1) \\ z(c^*) \leqslant z(c^*+1) \end{cases}$$

化简，得

$$L_s(c^*) - L_s(c^*+1) \leqslant \frac{c_s}{c_w} \leqslant L_s(c^*-1) - L_s(c^*)$$

依次计算序列 $L_s(1), L_s(2), L_s(3), \cdots$ 相邻两项之差，看常数 c_s/c_w 落在哪二者之间，便可确定出 c^*。

【例 8.17】某质检中心为各集团公司提供服务，做检验的公司的到达服从泊松流，平均到达率为每天 48 次，每次来检验由于停工等原因的损失为 600 元。服务时间服从负指数分布，平均服务率为每天 25 次，每设置 1 个检验员的服务成本(工资加设备折旧)为 400 元，其他条件同标准的 $M/M/c$ 系统。问应设几个检验人员可使总费用的平均值为最小。

北京大学出版社版权所有 禁止转载

解：据题知，$\lambda=48$，$\mu=25$，$\rho=\lambda/\mu=1.92$，$c_s=400$，$c_w=600$。设检验人员数量为c，依次令$c=1,2,3,4,5$按照下面公式计算L_s，其结果见表 8-4。

$$P_0=\left[\sum_{k=0}^{c-1}\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k+\frac{1}{c!(1-\rho)}\left(\frac{\lambda}{\mu}\right)^c\right]^{-1}\qquad L_s=\frac{(c\rho)^c\rho}{c!(1-\rho)^2}P_0+\frac{\lambda}{\mu}$$

表 8-4　计算过程表

c	1	2	3	4	5
$\lambda/c\mu$	1.92	0.96	0.64	0.48	0.38
L_s		21.61	2.68	2.068	1.952
$L_s(c)-L_s(c+1)$			18.93	0.612	0.116

依次计算L_s相邻两项之差，见表 8-4。由于$c_s/c_w=0.66$，其值落在 18.93 和 0.612 之间，所以取$c^*=3$，即安排 3 个检验人员可使总费用最小，总费用$z^*=2\,787$元。

8.5　随机模拟技术

在实际生产和生活中，有很多排队系统不能用标准模型描述，对于极简单的系统，可以用解析的方法解决，对于稍微复杂的系统特别是到达间隔时间和服务时间都不服从负指数分布的系统，解析法几乎无能为力，此时随机模拟技术就显示出了它的独特优势。

8.5.1　随机模拟技术的原理

模拟原指对客观事物的模仿，随着技术的发展和需要，人们原始的模仿行为就演化为系统的技术手段和方法。模拟又称为仿真(Simulation)，Korn 在 1978 年将仿真定义为“用能代表所研究系统的模型做实验”，1984 年 Oren 在给出了仿真的基本概念框架“建模—实验—分析”的基础上，指出了“仿真是一种基于模型的活动”的内涵，被认为是现代仿真技术的一个里程碑。仿真是根据所研究系统的结构组成、运行规则和运行过程，建立与之相似的模型，用一定的方法来模仿系统，以达到分析、研究和改善系统的目的。

模型可以是实物模型(如作战或建筑布局用的沙盘)或数学模型。对实物模型进行的模拟称为物理模拟，对数学模型进行的模拟称为数字模拟。物理模拟直观、形象，但模型的结构和参数改变困难、实验限制多、投资大。数字模拟改变系统结构和参数容易，灵活多变，便于对系统进行多方案的对比分析，投资小。数字模拟的工具是计算机，又称为计算机模拟，随着计算机的发展，计算机模拟逐步占据了主导地位。

下面仅对离散事件的数字模拟的原理(基于排队系统)做扼要的介绍。模拟过程主要是根据所研究系统的结构组成、运行规则和运行过程，建立与之相似的数学模型，再编写计算机程序或代码转化成计算机模拟模型，在计算机上调试运行，以获得所需结果，如图 8.14 所示。

模拟过程中有两类关键技术：一是随机变量的产生；二是信息的收集统计。

1. 随机变量的产生

在随机模拟中，必然有一些随机事件出现，模拟中需要按其概率规律来模仿这些随机

北京大学出版社版权所有 禁止转载

事件的产生。前提是首先要产生 0～1 的均匀随机数 u，然后再按一定的方法转化成符合分布规律的随机变量。比如顾客的到达间隔时间或服务时间服从某种概率分布，其分布函数为 $F(t)$，由于 $F(t)$ 的取值范围是 0～1，所以令 $F(t)=u$，利用这个函数关系应用“逆变换法”得到符合其概率分布的随机变量 T 的一个样本值 t，即

$$t=F^{-1}(u) \tag{8-12}$$

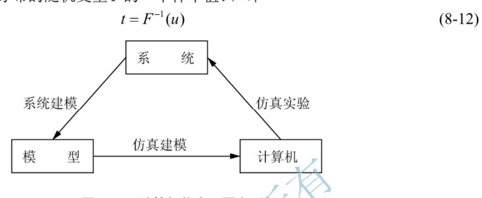


图 8.14 计算机仿真三要素

“逆变换法”仅适用于分布函数 $F(t)$ 存在显式的解析表达式情况，若不存在显式的解析表达式时，可用“舍选法”“组合法”“卷积法”和“变换法”等来产生符合其概率分布的随机变量 t，具体产生方法参见有关系统模拟的资料。

2. 信息的收集统计

系统模拟的目的是为了分析、研究和改善系统，因此模拟过程中必须根据需要收集相应的数据信息，以便模拟结束时统计计算得到系统的特征指标和有用信息。

蒙特卡罗方法，就是经典的随机模拟方法，这一方法源于美国在第二次世界大战研制原子弹的“曼哈顿计划”。该计划的主持人之一数学家冯·诺伊曼用驰名世界的赌城——摩纳哥的 Monte Carlo 来命名这种方法。蒙特卡罗方法的基本思想很早以前就被人们所发现和利用，早在 17 世纪，蒲丰(Buffon)提出著名的 Buffon 投针试验来近似计算圆周率 π 。

随着 20 世纪 40 年代电子计算机的出现，特别是近年来高速电子计算机的出现，使得人们在计算机上利用数学方法大量、快速地模拟这样的试验成为可能。计算机模拟程序可以用计算机高级语言或汇编语言编写，也可以用市场上的仿真专用商业软件来模拟，市场上的仿真商业软件较多，如 Witness、Flexsim、Arena、ProModel、VisFactory 等，这些仿真软件带有动画，模拟过程直观、逼真。目前这一方法已经广泛地被运用到数学、物理、管理、生物遗传、社会科学等领域，并显示出特别的优越性。

8.5.2 随机模拟技术的应用——随机服务系统

下面举例说明排队系统的随机模拟过程和应用。

【例 8.18】 某邮局接受顾客来邮寄包裹，设置 1 个服务台。顾客到达的间隔时间(分)服从 $U(2,7)$ 的均匀分布，服务时间为 4 分钟的固定时间，用数学模拟方法求顾客的平均等待时间和服务台的平均利用率。

解：由于顾客的到达时间服从均匀分布，根据分布函数

$$F(t)=\frac{t-a}{b-a} \qquad a\leqslant t\leqslant b$$

北京大学出版社版权所有 禁止转载

设 u 为 0～1 的均匀随机数，令 $F(t)=u$，应用“逆变换法”得到符合 $U(2,7)$ 概率分布的随机变量 t 为

$$t=F^{-1}(u)=a+(b-a)u=2+5u \qquad 0\leqslant u\leqslant 1 \tag{8-13}$$

为求顾客的平均等待时间和服务台的平均利用率，在模拟过程中就要收集每个顾客的等待时间和服务台的服务时间。下面对 10 个顾客的排队服务的模拟过程列于表 8-5。

表 8-5　10 个顾客的排队服务模拟过程

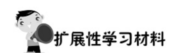
顾客到达顺序号	(0, 1)随机数	间隔时间	到达时间	当前时间	开始服务时间	服务结束时间	等待时间	服务时间
1	0.56	4.8	4.8	4.8	4.8	8.8	0	4
2	0.23	3.15	7.95	7.95	8.8	12.8	0.85	4
3	0.88	6.4	14.35	12.8	14.35	18.35	0	4
4	0.61	5.05	19.4	18.35	19.4	23.4	0	4
5	0.12	2.6	22	22	23.4	27.4	1.4	4
6	0.94	6.7	28.7	27.4	28.7	32.7	0	4
7	0.36	3.8	32.5	32.5	32.7	36.7	0.2	4
8	0.77	5.85	38.35	36.7	38.35	42.35	0	4
9	0.59	4.95	43.3	42.35	43.3	47.3	0	4
10	0.08	2.4	45.7	45.7	47.3	51.3	1.6	4

北京大学出版社版权所有 禁止转载

模拟从 0 时刻开始，产生第一个随机数为 0.56，代入公式(8-13)计算得第一个顾客的到达间隔时间是 4.8 分，即第 1 个顾客在 4.8 分时刻到达，此时服务台是空闲的，无须排队直接接受服务，服务时间是固定时间 4 分，在 8.8 分时刻结束。接着产生第二个随机数为 0.23，得第 2 个顾客的到达间隔时间是 3.15 分，应在 7.95 分时刻到达，由于此时服务台正在忙，他便参加排队等待。在 8.8 分时刻第 1 个顾客服务结束离开，排队等待的第 2 个顾客开始接受服务，记录第 2 个顾客的等待时间是 0.85 分，服务时间是固定时间 4 分，在 12.8 分时刻结束，接着是第 3 个、第 4 个顾客……以此类推，直到第 10 个顾客服务结束时刻为 51.3 分。

模拟结束后，按收集的数据统计计算所需结果，10 个顾客的等待时间总和为 4.05 分，所以顾客的平均等待时间为 $\hat{W}_q=4.05/10=0.405$ (分)，在此期间，服务台的服务总时间是 40 分，而系统运行了 51.3 分，因此服务台的平均利用率为 $\hat{S}_r=(40/51.3)\times 100\%=78\%$。

上面过程仅模拟了 10 个顾客，其模拟结果与理论结果相差是很大的，模拟的顾客数越多，结果才越接近理论值。利用 WinQSB 软件模拟 100 000 个顾客后得到顾客的平均等待时间为 $\hat{W}_q=1.442$ 分，服务台的平均利用率为 $\hat{S}_r=88.95\%$。W_q 和 S_r 的头上带有“^”符号，表明它们是估计值，而不是理论值。模拟运行一次，相当于做了一次随机试验，仅仅是一个样本，要得到可信的结果，要进行多次模拟运行得到很多样本，再求其均值，或进行区间估计，然后才能应用于决策。



Jackson 排队网络

排队网络包含一组节点，每个节点有若干服务器。单个节点可以被视为一个排队系统。客户可以在从任何节点进入排队网络。当客户在某个节点排队获得服务以后，它们可以离开网络，也可以去另外的节点，甚至回到原来的节点。

Jackson 排队网络满足：①一个排队网络有 k 个节点，节点 i ($i=1,2,\cdots,k$)上有 c_i 个服务器，服务器服务单个客户的时间服从指数分布，平均为 $1/\mu_i$，单个节点可以被视为一个 $M/M/c/\infty$ 排队系统；②客户到达是泊松过程，以速率 γ_i 到达节点 i；③当一个客户在节点 i 的服务器上完成服务，它以概率 r_{ij} 进入节点 j，以概率 r_{i0} 离开排队网络。对任意 i，若 $\gamma_i=0$，$r_{i0}=0$，称为闭合 Jackson 网络(没有客户进入和离开)，否则称为开放的 Jackson 网络。

Jackson 排队网络在计算机通信服务中具有重要的应用，如 800 免费电话服务就是典型 Jackson 排队网络：当用户拨号，在获得语音提示以后，可以按“1”进入产品介绍，按“2”进入问题咨询。用户用于听取语音提示并决定按“1”还是按“2”的时间平均为 30 秒，指数分布。一次只能有一名用户听取语音提示。大约 55%的用户选择产品介绍，产品介绍由三名客服人员负责，产品介绍平均 6 分钟，指数分布；45%的用户选择问题咨询，有 7 名客服人员负责问题咨询，问题咨询平均耗时 20 分钟，指数分布。大约有 2%的用户在听完产品介绍以后转到问题咨询，1%的用户在咨询完问题以后转到产品介绍。每小时平均有 35 名用户拨打电话，泊松过程。当用户不能获得服务时，他们在电话里听到音乐并且等待。

Jackson 排队网络理论也能应用于港口集装箱码头的物流网络进行重新设计，对港口局部物流网络进行了诊断、优化，可大大提高了港口的装卸效率、效益与服务水平。

北京大学出版社版权所有 禁止转载

本 章 小 结

排队系统在社会生活中应用广泛，其主要由到达顾客、等待的队列，以及提供服务的服务机构组成。本章从排队系统的输入过程、排队规则和服务机构三要素引入，首先学习了排队系统的分类规则；其次对负指数分布的单服务台和多服务台排队系统进行了详细的绩效考核；然后过渡到一般服务时间分布排队系统的探讨分析，并就简单的排队系统进行了优化设计；最后给出了随机服务系统的模拟的蒙特卡罗法，为复杂的排队系统的研究展现了新的研究思路。



关键术语(中英文)

排队论(Queuing Theory)
泊松过程(Poisson Process)
负指数分布(Negative Exponential Distribution)
爱尔朗分布(Erlang Distribution)
排队规则(Queue Discipline)
绩效指标(Performing Index)
服务强度(Traffic Intensity)
平均队长(Expected Value of Team Length)
平均队列长(Expected Value of Queue Length)
平均逗留时间(Expected Value of Sojourn Time)

平均等待时间(Expected Value of Waiting Time)　　忙期(Busy Period)
随机模拟技术(Random Simulation Method)　　蒙特卡罗法(Monte Carlo Method)



知识链接

排队论——爱尔朗

Agner Krarup Erlang (1878—1929)，是研究电话网络问题的第一人，他是哥本哈根电话公司的工程师，发现了 Erlang B、Erlang C 及其他一些话务量公式，在金融、电信、运输、网络、呼叫中心等领域具有广泛的运用。



爱尔朗 1878 年出生在丹麦，早期在他父亲管理的学校里度过，他的父亲是村校长和教区办事员。他最喜欢的科目是天文学，小学教育完成后他在哥本哈根接触到了数学和自然科学，16 岁时获得了哥本哈根大学奖学金并于 1901 年完成了他的学业。在接下来的 7 年里，他在多个学校讲学，教学质量优良，受到学生的好评。他不善交际，善于观察，养成了简洁的讲话风格。他经常利用假期到国外旅行，去法国、瑞典、德国和英国参观艺术画廊和图书馆。

在教学的同时，爱尔朗仍不懈地坚持数学和自然科学方面的研究。他是丹麦数学家协会的成员，哥本哈根电话公司的会员。他于 1908 年到这家公司工作，后来担任公司实验室的负责人。在研究电话通信系统时，爱尔朗基于描述某一随机现象在一定时间内重复概率的“泊松分布”，根据用户、电话线路以及接线员的数量，计算在哥本哈根电话网中从拨号到接通所需的时间，研究人们打电话的行为模式，推导出人们需等待多久的著名的排队公式，并于 1909 年发表了关于排队理论的第一篇论文。

北京大学出版社版权所有 禁止转载



习题 8

8.1 填空题

(1) 排队系统中 $X/Y/Z/A/B/C$，符号“B”表示__________。

(2) 排队模型 $M/M/2$ 中的 M、M、2 分别表示到达间隔时间为__________分布，服务时间服从负指数分布和服务台数为 2。

(3) 某企业有 5 台运货车，已知每台车每运行 100 小时平均需维修两次，一个维修工，每次需时 20 分钟，到达时间和服务时间均服从负指数分布，该问题的排队模型属于__________。

(4) 顾客输入为泊松流参数为 λ，服务时间为负指数分布参数为 μ，一个服务站，平均服务时间为 6 秒，每分钟有 5 名顾客到达。则 $\lambda=$__________； $\mu=$__________。

(5) 泊松流满足条件__________、__________、__________。

8.2 判断题

(1) 如果顾客连续接受串联的几个服务台的服务，各服务台服务时间相互独立，且服从参数为 μ 的负指数分布，那么顾客接受几个服务台总共所需时间服从爱尔朗分布。 (　　)

(2) 某寻呼台有两个接线员，平均每分钟有 4 人打进电话，平均每个客户服务时间为 0.4 分钟，该寻呼台话务强度为 0.8。 (　　)

(3) 排队系统中，顾客等待时间的分布不受排队规则的影响。 (　　)

(4) 在排队系统中，多队多服务台与单队多服务台相比，其服务台的平均利用率是相同的。 ( )

(5) 若到达排队系统的顾客为泊松流，则依次到达的两名顾客之间的间隔时间服从负指数分布。 ( )

8.3 指出下列排队系统中的顾客和服务员：(1)自行车修理店；(2)按客户订单进行加工的加工车间；(3)机场起飞的客机；(4)十字路口红灯前的车辆。

8.4 表 8-6 中的数据是到达邮局的顾客数和对顾客服务时间的统计结果。以 3 分钟为一个时段，统计了 100 个时段中顾客到达的人数以及对 100 位顾客的服务时间，见表 8-6。若已知顾客到达间隔时间和服务时间服从负指数分布，试用统计方法确定该邮局的平均到达率 λ 与平均服务率 μ 。

表 8-6 到达邮局的顾客数和对顾客服务时间的统计结果

到达人数	0	1	2	3	4	5	6
时段数	14	27	27	18	9	4	1
服务时间/秒	0～12	12～24	24～36	36～48	48～60	60～72	72～84
顾客人数	33	22	15	10	6	4	3
服务时间/秒	84～96	96～108	108～120	120～150	150～180	180～200	>200
顾客人数	2	1	1	1	1	1	0

北京大学出版社版权所有 禁止转载

8.5 对 $M/M/1/\infty/\infty$ 的排队系统，

请根据下列表达式分别解释其含义：(1) λ/μ ；(2) $P\{i>0\}$ ；(3) $L_s - L_q$ ；(4) $W_s - W_q$ 。

8.6 某理发馆有两名理发师，另放 5 把椅子供顾客等待，当座位全部坐满时，后来者便自动离去。顾客到达间隔与理发时间均为相互独立的负指数分布，每小时达 3.76 人，每人理发平均需要 15 分钟。试求潜在顾客的损失率及平均逗留时间。

8.7 设有两个售票亭，现考虑每分钟平均到达 6.4 人的泊松流，服务时间服从负指数分布，平均每分钟可服务 4 人。试求系统中无人的概率，系统中的平均人数，排队等候的平均人数，顾客等候的平均时间。

8.8 某机场有两条跑道，每条跑道只能供一架飞机降落，平均降落时间为 2 分钟，并假定飞机在空中等待的时间不得超过 10 分钟，试问该机场最多能同时接受多少飞机降落?

8.9 某电信局准备在新建成的国际机场装设电话亭，而电信局的目标是每一个等候电话的概率不超过 0.10；使用电话的平均需求率为每小时 30 人，且为泊松流，使用电话的平均时间为 5 分钟，且为负指数分布。应该设置多少个电话亭？

8.10 某大型预制厂在考虑卸货台装卸材料时需对甲、乙、丙三个方案作出决策。设 F、c_e、μ 分别表示各方案中每天的固定费用，每天的可变费用，每小时的装卸 a 吨。三方案的有关费用资料见表 8-7。再设货车按泊松流到达，每天平均(按 10 小时计算)到达 150 车，每车平均装货 5 吨。卸车(货)时间服从负指数分布，每辆车停留在装卸站 1 小时的损失费为 10 元，问应采取什么方案才能使所花费的总费用为最小？

表 8-7　三方案的费用资料

方　　案	F	c_e	a	u
甲	60	10	100	20
乙	130	15	200	40
丙	250	20	600	120

8.11 某厂机修车间专门修理某种大量的设备。该设备故障数为泊松流，平均每天 2 台，修复时间服从负指数分布，平均每台的修理时间为$1/\mu$天。μ与机修人员及维修装备现代化程度(即与机修车间全年经费K元)等有关，为

$$\mu = 0.1 + 0.001K \quad (K \geqslant 1900)$$

每台故障设备每天的停产损失为 400 元。试确定该厂最经济的机修年经费K及平均修复率μ。

8.12 已知某材料订货后，到货延误天数的随机数分布见表 8-8。若由随机数表依次查得随机数为 15、09、41、74、00、72、67、55、71、35，试给出今后 10 次订货实际到货时间被延误的模拟样本。

表 8-8　到货延误天数的随机分布

到货延误天数	概　　率	随机数分布
5	0.03	00～02
4	0.09	03～11
3	0.23	12～34
2	0.38	35～72
1	0.20	73～92
0	0.07	93～99

北京大学出版社版权所有　禁止转载



实际操作训练

某物流公司有一个轮船装卸码头，货船的到达为泊松过程，由于经济的发展，到达率由以前的 2 艘/天提高为现在的 3.6 艘/天。码头有 1 个装卸泊位和 1 个供等待的泊位，装卸时间服从负指数分布，服务率为 3 艘/天。每个泊位的单位费用为C_r=2 000 元/天，单船服务费用是函数$C_s = 8\,000[1 + 0.4(c-1)]$，其中$c$是装卸泊位数量。每装卸一只轮船的收入为$G = 18\,000$元。显然当前系统的损失率很大(服务率小于到达率)，公司管理者准备改善码头的服务设施，拟定了 5 个方案，分别是：

方案 1，再新增 1 个等待泊位，其他条件不变；

方案 2，改善装卸设施，提高服务率到 4 艘/天，其他条件不变；

方案 3，将原等待泊位改造成装卸泊位，即码头有 2 个装卸、0 个等待泊位，其他条件不变；

方案 4，再新增 1 个装卸泊位，即码头有 2 个装卸、1 个等待泊位，其他条件不变；

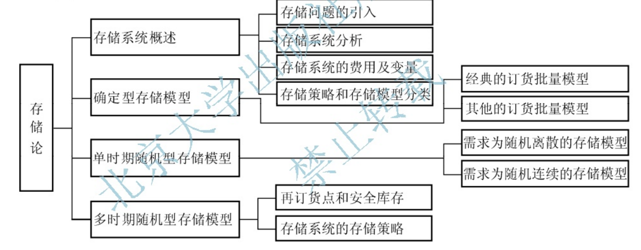
方案 5，改造和新增到 3 个装卸泊位，0 个等待泊位，其他条件不变。

问哪个方案可使物流公司每天的利润最高？

第 9 章 存 储 论



【本章知识架构】



北京大学出版社版权所有
禁止转载

【本章教学目标与要求】

- 掌握存储系统的结构，熟悉存储系统的运营费用、存储策略和存储模型的分类。
- 掌握确定型存储模型的费用构成，熟悉各种确定型存储模型，会用其解决相应的问题。
- 掌握单时期随机型存储模型在管理活动中的应用。
- 理解存储系统再订货点和安全库存的含义，了解常见的几种存储策略在多时期随机型存储模型中的应用。



运储物流公司的库存管理

仓储成本约占物流总成本的三分之一，仓储管理的目标之一就是实现最低库存，减少资产负担和提高相关的周转速度。许多企业采取多种管理措施控制库存：批量订货、联合库存管理、供应商管理库存等，这些策略可以提高制造商管理的效率，减少商品缺货率，减少动作成本，提高对市场变化的反应速度，更好地满足消费者的需求。

作为第三方物流的运储物流公司，通过专业化和网络化的运作，将仓储管理打造成控制中心、调度中心、增值服务中心，帮助服务商实施以下内容。

（1）灵活地配送任何数量的产品，分摊车辆配送的固定成本，降低服务商的单位库存成本，在配送过程以“零库存”为理论目标。

（2）更好地满足制造业、零售业、电商业等不同行业的个性化需求，其建立的仓储物流中心可以实现订单管理、仓储、包装和配送等多种功能，在国内一些中心城市也能实现次日达的配送服务。

（3）通过智能化的信息集成，满足了各种精益化、数字化及网络化的运营需求，建成了高效方便的云信息管理平台，强化了与服务商之间的精诚合作。

北京大学出版社版权所有 禁止转载

存储论又称库存理论，是运筹学中发展较早的分支，早在 1915 年，哈李斯(F. Harris)针对银行货币的储备问题进行了详细的研究。到了 20 世纪 50 年代，人们开始应用系统理论来研究和解决库存问题，存储论成了运筹学的一个独立分支。存储是系统随机聚散现象，存储的作用在于缓冲调节供求之间的不平衡，以避免由需求大于供应而造成的损失；但存储也有损失，需要支付存储费用。

存储论研究的基本问题是，对于特定的需求类型，讨论用怎样的方式进行物质的供应、商品的订货或者产品的生产，以求最好地实现存储的经济管理目标。具体地说，存储论是研究如何根据生产或者销售活动的实际存储问题建立起数学模型，然后通过费用分析求出产品、商品的最佳供应量(量)和供应周期(期)这些数量指标。

9.1 存储系统概述

库存(Inventory or Stock)表示用于将来目的的资源暂时处于闲置状态，库存的存在主要由供需双方在时间、空间和数量上的不确定性所引起。存储管理是物流活动的重中之重，丰田生产方式认为：“库存是万恶之源”，强化库存量的控制可以改善企业经营水平，达到减少资金占用，获得更多利润的目的。这里先给出存储理论的一些基本概念。

9.1.1 存储问题的引入

现代化的生产和经营活动都离不开存储，生产实践中由于种种原因，需求与供应、消费与存储之间存在着不协调性，其结果将会产生两种情况，一种情况是供过于求，由于材料、产品或者商品的积压，造成资金周转的缓慢和成本的提高而带来经济损失。例如，一个生产企业为了连续进行生产，就需要储备一定数量的原材料或半成品。另一种情况是供

不应求，由于原料或者商品短缺，引起生产停工或者无货销售，使经营单位因利润降低而带来经济损失，例如一个超市为了满足顾客的需求，就必须有足够的商品库存。因此，存储问题是生产活动中一个普遍存在的问题。物资的存储，除了用来支持日常生产经营活动外，有库存的调节还可以满足高于平均水平的需求，同时也可以防止低于平均水平的供给。此外，有时大批量物资的订货或利用物资季节性价格的波动，可以得到价格上的优惠。为了使经营活动的经济损失达到最小或者收益实现最大，人们开始关注在供应和需求之间的存储环节，开始研究如何寻求原料、产品或者商品合理的存储量以及它们合适的存储时间，来协调供应和需求的关系，这正是存储论所要研究的问题。

【例 9.1】 某商店出售某商品，每单位 A 商品的成本 500 元，顾客对该商品的年平均需求量为 365 单位，需求率为常数(即每天需求 A 商品 1 单位)，不允许缺货。如该店每次订货都能得到及时供给，问该店每隔多少天订一次货？每次订货量为多少？才能使年内总费用支出最少(费用支出包括订货费、商品在店内的存储费)。已知该店存储 A 商品每单位每年的费用为商品成本的 20%，每次订购商品需 20 元的订货费。

解：这里仅分析两种情况。

方案 1：每两天订货一次，每次订购 2 单位。此情形下库存量在 0～2 变化，平均库存量显然为每天 1 单位。

每年库存费用(500×20%×1)元＝100 元；每年订货次数 365÷2＝182.5 次

每年订货费用(182.5×20)元＝3 650 元；合计总费用 100+3 650＝3 750(元)

方案 2：每 20 天订货一次，每次订购 20 单位。此情形下库存量在 0～20 变化，平均库存量显然为每天 10 单位。

每年库存费用(500×20%×10)元＝1 000 元；每年订货次数：365÷20＝18.25 次

每年订货费用(18.25×20)元＝365 元；合计总费用：(1 000+365)元＝1 365 元

比较方案 1 和方案 2，可知方案 1 的库存费用低，但订货费用高；方案 2 的库存费用高，但订货费用低，从总费用来看，方案 2 优于方案 1。

可以预知，如继续设计适当的方案，有望能使总费用继续降低，如每 10 天订货一次，每次订购 10 单位时，则总费用为 1 230 元。问题的实际背景也提示我们，各种方案的不同总费用肯定为正数，一定存在一个方案其总费用是最少的。原则上，可以用如上的方法试探求出充分接近这一使总费用达到最少的方案(即求出订货间隔与订货量)。然而，实际上几乎不可能这样做，一是因为计算量庞大；二是因为不能或者说不容易求得使总费用刚好达到最小的订货间隔与订货量的精确值。

存储理论的目的之一就是为迅速而精确地求出以上问题的解，应用数学模型化的方法，给出解决问题的决策支持。它所采用的一般方法是首先分析存储系统的要素和其结构，其次是建立相应的数学模型，最后求解数学模型得到精确的或满意的解。

9.1.2　存储系统分析

存储的最终目的就是协调供求关系，为了考察最佳的库存方案，这里首先分析存储系统包含哪些要素(环节)。物资的存储，包括生产存储，为了维持企业正常生产而储备的原材料或半成品；产品存储，为了满足其他部门的需要而存储的半成品或成品；供销存储，为了满足顾客的需要，存储在供销部门的各种物资。不论哪种存储系统，一般都可用图 9.1 的模式来表示。

北京大学出版社版权所有 禁止转载

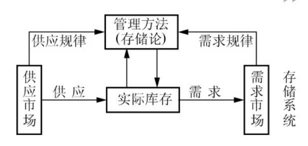


图 9.1　存储系统及其管理

存储系统可以用“供—存—销”三个字来描述，即一个存储系统，通过订货以及进货后的存储与销售来满足顾客的需求。

1. 存储系统的输出

从存储系统中取出一定数量的库存货物，这就是存储系统的输出。存储系统的输出方式有间断式和连续式，如图 9.2 所示，其中 S 是初始库存量，经过时间 T 后，库存量是 W，输出了 $S-W$。

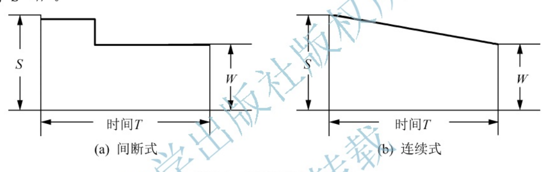


图 9.2　存储系统输出模式

北京大学出版社版权所有
禁止转载

需求方每次提出的需求量可能是确定的，也可能是随机的。如某工厂每月需要钢材 10 吨是固定不变的，而对机器零件的需要量却是每月都在变，如 1 月需要 40 个、2 月需要 55 个、3 月是 30 个等，一般根据大量的统计数据，找出需求量满足的统计规律。

2. 存储系统的输入

本章默认凡待使用的或待销售的物资即是进入了库存。库存物资由于不断地输出而减少，必须进行及时补充，库存的补充就是存储系统的输入，它可以通过订货、采购等活动来进行，也可以通过内部的生产活动来进行。

正如我们熟知的，一个商店为了提高顾客选购商品的满意度，就必须储存商品。这样它首先要掌握进货渠道，即供货市场，供货市场的变化规律和行情是它所关心的。其次它还应掌握需求市场，即顾客对商品的需求规律、需求方式等。清楚了供应和需求的变化规律，商店才可根据自己的情况，应用库存论的方法来控制实际库存。存储系统的运行规律主要是由供应市场与需求市场的运行规律所决定的，而实际中供应与需求的规律是多种多样的。这些不同的规律，显然从根本上决定了库存系统的运行规律。

在存储系统中对库存进行管理是必要的，它在整个供应—生产—分配过程中起着缓冲的作用。由于需求量往往是系统外部决定的，因而存储系统的输出难以控制和掌握，而影响存储系统输入的很多因素则可以自己来控制，这些因素主要是补充库存的时间以及补充的数量这些参数，以及为保持库存，需要进行订货，开始订货到进货有一段时间，因此为

保证及时供应，就要提前订货，提前的时间称之为“提前订货时间”，提前订货时间可以是确定的，也可以是随机的。

9.1.3　存储系统的费用及变量

系统的决策者可以通过控制订货时间的间隔和订货量的多少来调节系统的运行，使得在某种准则下系统运行达到最优。为此要明确该系统所发生的费用和有关的决策变量。

1. 存储系统的费用

费用是存储管理的一个重要经济指标，存储系统必须按最经济的原则运行，为了建立存储模型必须了解各类存储系统费用的构成情况。

1) 订货费

对供销型存储系统来说，订货费是指为补充库存，办理一次订货发生的有关费用，包括订货过程中发生的订购手续费、联络通信费、人工核对费、差旅费、货物检查费、入库验收费等。

当生产型存储系统自行组织生产时，订货费相当于组织一次生产所必需的工具夹安装、设备调整、试车、材料安排等费用。订货费一般说来与订购或生产的数量无关或基本无关。在确定订货费时，对具体问题要具体分析，但必须注意不能将搬运费、管理费等平均分摊到每一件货物上去。这样，就使订货费和一次订购的数量有关了。在年消耗量固定不变的情况下，一次订货量越大，订货次数就越少，订货费就越少。因此，从订货费角度看，订货批量越大越好。

2) 存储费

一般是指库存物资每单位时间所需花费的费用，也称持有成本。这一项费用中，只计入与库存物资数量成正比的部分，凡与存储物资数量无关的不变费用不计算在内。有时存储费还经常用每存储 1 元物资单位时间所支付的费用来表示，称为存储费率。

存储费包括存储物资所占用资金的利息、物资的存储损耗、陈旧和跌价损失、存储物资的保险费、仓库建筑物及设备的修理折旧费、保险费、存储物资的保养费、库内搬运设备的动力费、搬运工人的工资等。在以上存储成分中，利息支出所占比重较大。以工业贷款月利率 0.6%计算，存储百万元物资一年，仅利息就需支付 7.2 万元。由此可见，控制存储物资数量，对加速物资周转的意义。

由于订货量越大，平均库存量就越大，从而存储费支出越大。因此，从存储费角度看，订货批量越大越不好。

3) 缺货费

在存储系统中，还有一种费用是存储系统管理者不愿发生的费用，这就是缺货费。这项费用是因为供需脱节造成生产停顿而进行计划调整的损失费，或加班突击所增加的额外开支，未能完成合同任务而使需方损失给供方带来的赔款损失费等，还可能是因为无货满足顾客而降低服务质量、损失信誉及潜在盈利机会，丧失顾客而造成销售不良的损失等。总之，这是由于缺货而未能满足需求所带来的各种损失的费用表现。在有些情况下是不允许缺货的。如战争中缺少军械、弹药等将造成人员重大伤亡乃至战败，血库缺血将造成生命危害等，这时的缺货费可视为无穷大。

4) 生产费(买价)

如果库存不足需要补充，可选外购或自行生产。外购时需支付买价(当有折扣时更要考

北京大学出版社版权所有 禁止转载

虑买价)；自行生产时，这里的生产费用专指与生产产品的数量有关的费用如直接材料、直接人工、变动的制造费用。

上述费用构成了存储系统中最重要的几种费用。实际问题中有时是很难将它们绝对分离开的，但这些都是应用中的技术问题，根据不同的情况才能进行详细的分析。在这里只是从原则上对它们进行大致的分类。尽管在实际存储管理中，可能还会涉及其他各种名目的费用，但一般情况下，存储问题的费用总归是由上述四种费用所组成，并且它们与存储总费用(Inventory Cost)有线性关系，且与下列一些存储系统中的变量有着密切联系。

2. *存储系统的变量*

在存储系统中经常分析以下变量。

1) 订货批量 Q

存储系统根据需求，为补充某种物资的存储量而向供货厂商一次订货或采购的数量。

2) 订货周期 T

两次订货的时间间隔或订货合同中规定的两次进货之间的时间间隔。

3) 报警点 s

又称订货点，该点库存量和提前订货时间是相对应的，当库存量下降到这一点时，必须立即订货，当所订的货物尚未到达及入库之前，存储量应能按既定的服务水平满足提前订货时间的需求。

4) 安全库存量 ss

又称保险储备量。由于需求量 D 和提前订货时间 t 都可能是随机变量，随其波动幅度的增大可能大大超过其平均值，为了预防和减少这种随机性造成的缺货。必须准备一部分库存，这部分库存称为安全库存量。只有出现缺货情况时才动用安全库存量。

5) 最高库存量 S

在提前订货时间可以忽略不计的存储模型中，S 指每次到货后所达到的库存量。当存在提前订货时，S 指发出订货要求后，库存应该达到的数量，由于此时并未实际到货，所以该最高库存量又称名义库存量。

6) 最低库存量

一般是指实际的库存最低数量。

7) 平均库存量 $\overline{Q}$

库存保有的平均库存量，当存在报警点 s 时，平均库存量为

$$\overline{Q}=\frac{1}{2}Q+s$$

8) 记账间隔期 R

指库存记账制度中的间断记账所规定的时间，即每隔 R 时间，整理平时积欠下来的发料原始凭据，进行记账，得到账面结存数以检查库存量。

9.1.4 存储策略和存储模型分类

在存储系统中，将物质保持在预期的一定水平，使得生产和流通过程不间断并有效地进行，需对输入过程中的订货时间和订货数量进行控制，称为存储策略。存储策略确定后，首先是把实际问题抽象为数学模型，对一些复杂的条件尽量加以简化，只要模型能反映问

北京大学出版社版权所有 禁止转载

题的本质就可以了；然后用数学的方法对模型进行求解，得出数量的结论；结论是否正确，还要到实践中加以检验，如结论不符合实际，则还要对模型加以修改，重新建立、求解、检验，直到满意为止。

1. 常用存储策略

存储系统策略包含定量订购和定期订购两大类。

1) 定量订购制

泛指通过公式计算或经验求得报警点s和每次Q，并且每当库存量下降到s点时，就进行订货的存储策略，通常使用的有(Q,s)、(S,s)、(R,S,s)制等。

(1) (Q,s)制。采用这种策略需要确定订货批量Q和报警点s两个参数，(Q,s)属于连续监控制(又称永续盘点制)，即每供应一次就结算一次账，得出一个新的账面数字并和报警点s进行比较，当库存量达到s时，就立即以Q进行订货。

(2) (S,s)制。这种策略是(Q,s)制的改进，需要确定最高库存量S及报警点s两个参数。(S,s)制属于连续监控制。每当库存量达到或低于s时，就立即订货，使订货后的名义库存量达到S。因此，每次订货的数量Q是不固定的。

(3) (R,S,s)制。这种策略需要确定记账间隔期R、最高库存S和报警点s三个参数。(R,S,s)制属于间隔监控制，即每隔R时间整理账面，检查库存，当库存等于或低于s时，应立即订货，使订货后名义库存量达到S，因而每次实际订购批量是不同的，当检查实际库存量高于s时，不采取订货措施。

2) 定期订购制

即每经过一段固定的时间间隔T(称订货周期)就补充订货使存储量达到某种水平的存储策略，常用的有(T,S)制。

(T,S)制需要确定订购间隔期T和最高库存S两个参数，即每隔时间T检查库存，根据剩余存储量和估计的需求量确定订货量Q，使库存量恢复到最高库存S。

2. 存储模型分类

为便于讨论，根据不同的供求规律、库存产品的种数和考虑的时间周期对存储系统进行分类。

1) 确定型与随机型存储模型

凡需求量D、提前订货时间t为确定已知的存储问题所构成的存储模型为确定型；凡上述二者之一或全部为随机变量的存储问题构成的存储模型为随机型。

这种分类是下面最经常用的分类方法，其他分类方法用得不多。例如，商店经销某种日用品，该日用品的需求量服从某一随机分布规律，则该日用品的存储模型就是随机型的；又如修路需某种型号的水泥，若每日需求量基本上是固定的，供货水泥厂货源充足，用料单位组织进料运输，因此可以认为需求量、提前订货时间均为确定已知的，该种水泥的存储模型就是确定型。

在确定型存储模型中，又可分为需求不随时间变化和需求随时间变化两种类型；同样，随机型存储模型也可根据需求量是否随时间变化分为两类。实际中所谓绝对的确定型是不存在的，D,t多多少少总会有一些波动的，一般若随机变量x的均值为$\bar{x}$，标准差为σ_x，只要变异系数$c_x=\sigma_x/\bar{x}$小于0.2，随机变量x就可以当作确定型变量来对待。

北京大学出版社版权所有 禁止转载



2) 单周期与多周期存储模型

有的物资必须购进后一次全部供应或售出，否则就会造成经济损失，这类存储问题的模型称为单周期存储模型，如报纸、年历等时令性物品以及防洪、防冻季节性物资构成的模型。有的物资多次进货多次供应，形成进货—供应消耗—再进货—再供应消耗，周而复始的多周期特点的存储问题的模型称为多周期存储模型。

3) 单品种与多品种存储模型

一般地，将数量大、体积大又占用金额多的物资单独设库管理，称为单品种库。如木材、水泥、焦炭、煤等，这类库存往往占用大量资金，要采用比较精细的方法来计算其存储控制参数。有些物资是多品种存放在一个仓库里，就称为多品种库，如钢材库、电器元件库、配件库、有色金属库等。多品种库的存储不可能逐一计算每种物资的库存控制参数，可以将库存物资按其占用金额进行 *ABC* 分类存储管理。由于流动资金定额一般是按仓库下达的，所以多个品种物资存放在一个仓库时，往往存在资金约束及仓库容积约束，这样的存储模型称为带约束的存储问题。本章着重介绍单品种存储模型，多品种存储问题可同样讨论。

在存储模型中，目标函数是选择最优策略的准则。常见的目标函数是关于总费用或平均费用或折扣费用(或利润)的，最优策略的选择应使费用最小或利润最大。综上所述，一个存储系统的完整描述需要知道需求、供货滞后时间、缺货处理方式、费用结构、目标函数以及所采用的存储策略、决策者通过何时订货、订多少货来对系统实施控制。

北京大学出版社版权所有 禁止转载

9.2 确定型存储模型

确定型存储模型是最简单的存储模型，这类模型的有关参数如需求量、提前订货时间是已知确定的值，而且在相当长一段时间内稳定不变。经过数学抽象概括的存储模型虽然不可能与现实完全等同，但对模型的探讨将加深对存储系统的认识，其模型的解也将对存储系统的决策提供帮助和依据。

9.2.1 经典的订货批量模型

经典的经济订货批量模型就是下面的模型 1，也称为 EOQ 模型。

1. 模型 1：不允许缺货，生产时间很短

1) 模型假设

(1) 缺货费用无穷大。

(2) 当库存降至零时，可以立即得到补充(即生产时间很短，可以近似地看做零)。

(3) 需求是连续的、均匀的。

(4) 每次订货量不变，订购费不变(每次生产量不变，装配费不变)。

(5) 单位存储费不变。

这就是著名的经济订货批量 EOQ(Economic Order Quantity)模型假设。经济订货批量模型又称整批间隔进货模型，该模型适用于整批间隔进货、不允许缺货的存储问题，即某种物资单位时间的需求量为常数 D，存储量以单位时间消耗数量 D 的速度逐渐下降，经过时间 T 后，存储量下降到零。此时开始订货并随即到货，库存量由零上升为最高库存量 Q，然后开始下一个存储周期，形成多周期存储模型。

该模型的需求量和提前订货时间是确定的，只要确定每次订货的数量是多少或进货间隔期为多长，就可以作出存储策略。由于存储策略是使存储总费用最小的经济原则来确定订货批量的，故称该订货批量为经济订货批量。

2) 模型的有关参数

T——订货周期或存储周期(年或月或日)；

D——需求数量，即单位时间的需求量(件/年或件/月或件/日)；

Q——每次订货批量(件或个)；

c_1——存储单位物资单位时间的存储费[元/(件×年)或元/(件×月)或元/(件×日)]；

c_3——每次订货的订货费(元)；

K——货物单价(元/件)。

3) 模型建立和求解

模型的存储量变化如图 9.3 所示。一个订货周期内需要该种物资 $Q = DT$ 个，图中存储量斜线上的每一点表示在该时刻的库存水平，每一个订货周期存储量的变化形成一个直角三角形。

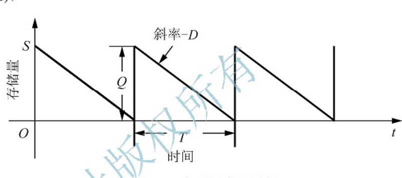


图 9.3 存储量变化过程

一个订货周期的平均存储量为 $\frac{1}{T}\int_0^T D t \mathrm{d}t = \frac{1}{2}Q$，存储费用为 $\frac{1}{2}c_1 QT$；一次订货费用为 c_3，一个存储周期的订货总费用 $c_3 + DKT$，因此，在这段存储周期内总费用为 $\frac{1}{2}c_1 QT + c_3 + DKT$。

北京大学出版社版权所有 禁止转载

由于订货周期 T 是变量，为便于比较，计算单位时间的平均总费用，即

$$C_Z = \frac{1}{2}c_1 Q + \frac{c_3}{T} + DK$$

将 $T = \frac{Q}{D}$ 代入，得

$$C_Z = \frac{1}{2}c_1 Q + \frac{c_3 D}{Q} + DK \tag{9-1}$$

显然，单位时间的订货费随着订货批量的增大而减小，而单位时间的存储费随订货批量 Q 的增大而增大，如图 9.4 所示。

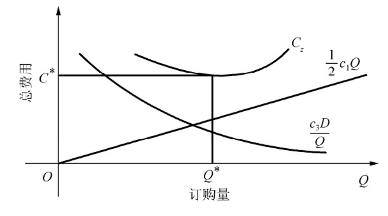


图 9.4 存储费用曲线

求导数

$$\frac{\mathrm{d}C_Z}{\mathrm{d}Q}=\frac{1}{2}c_1-\frac{c_3D}{Q^2}=0$$

即得经济订货批量

$$Q^*=\sqrt{\frac{2c_3D}{c_1}} \tag{9-2}$$

这是因为$\frac{\mathrm{d}^2C_Z}{\mathrm{d}Q^2}=\frac{2c_3D}{Q^2}>0$，故当$Q^*=\sqrt{\frac{2c_3D}{c_1}}$时，$C_Z$取最小值。

公式(9-2)称为经济订货批量公式，由于威尔逊是该公式推导应用的倡导者，所以该公式也称为威尔逊公式。

由式(9-2)及$Q^*=TD$可得到订货周期$T^*=\sqrt{\frac{2c_3}{Dc_1}}$，将$Q^*=\sqrt{\frac{2c_3D}{c_1}}$代入(9-1)式前一部分，立刻得到按经济订货批量进货时的最小存储总费用为

$$C^*=\sqrt{2Dc_1c_3} \tag{9-3}$$

由于Q^*，T^*均与K无关，所以此后的费用函数中省略去DK项。还要说明的是，前面在确定经济订货批量中，认为订货和进货可以同时发生，实际上，订货和到货一般总有一段时间间隔，为保证供应的连续性，需要提前订货。设提前订货时间为t，日需要量为D，则订购点$s=Dt$，当库存下降到s时，即按经济订货批量Q^*订货，在提前订货时间内，以每天D的速度消耗库存，当库存下降到零时，恰好收到订货，开始一个新的存储周期。

对于以上确定型存储问题，最常使用的策略就是经济订货批量Q^*，并每隔时间T^*即订货，使存储量恢复到最高库存量Q，这是实际中最常用的经济订货批量公式。该存储策略可以认为是定量订购制，但因订购周期也固定，又可以认为是定期订购制。

【例 9.2】 某企业每年对某种零件的需求量为 20 000 件，每次订货的固定订货费用为 1 000 元，该零件的单价为 30 元，每个零件每年的保管费用为 10 元，求经济订货批量，每年的库存总成本，每年的订货次数和每次订货的时间间隔(一年按 360 天计)。

解：据题知$D=20\,000$件、$c_1=10$元/件、$c_3=1\,000$元/次、$K=30$元/件，则

(1) 订货批量 $Q^*=\sqrt{\frac{2c_3D}{c_1}}=\sqrt{\frac{2\times1\,000\times20\,000}{10}}=2\,000$(件/次)；

(2) 每年的最小存储总费用$C^*=\sqrt{2c_1c_3D}=\sqrt{2\times10\times1\,000\times20\,000}=20\,000$(元)，

考虑购买成本的总费用为

$$TC=DK+C^*=(20\,000\times30+20\,000)\text{元}=620\,000\text{元；}$$

(3) 每年的购买次数 $N=\frac{D}{Q}=\frac{20\,000\text{件}}{2\,000\text{件/次}}=10$次；

(4) 每次订货的时间间隔 $T^*=\sqrt{\frac{2c_3}{c_1D}}=\sqrt{\frac{2\times1\,000}{10\times20\,000}}=\frac{1}{10}$年$=36$天。

若以D表示某种物资的年需用量，K表示该物资的单价，c_3为一次订货费，r表示存储费率，即存储每元物资一年所需的存储费用，则得到经济订货批量的另外一种常用形式$Q^*=\sqrt{\frac{2c_3D}{rK}}$。

北京大学出版社版权所有 禁止转载

2. 模型1的灵敏度分析

灵敏度分析主要分析模型中各参数变化时对订货批量和总成本的影响。

1) 需求量对经济订货批量及总成本的影响

设需求量增加了δ倍，变化后的需求量为$D'=(1+\delta)D$，则变化后的订货批量为

$$Q'=\sqrt{\frac{2c_3D'}{c_1}}=\sqrt{\frac{2c_3(1+\delta)D}{c_1}}=\sqrt{\frac{2c_3D}{c_1}}\sqrt{1+\delta}=Q^*\sqrt{1+\delta}$$

总费用为(不包括货物购买成本)：

$$C'=\sqrt{2D'c_1c_3}=\sqrt{2(1+\delta)Dc_1c_3}=\sqrt{2Dc_1c_3}\sqrt{1+\delta}=C^*\sqrt{1+\delta}$$

2) 订货批量对总成本的影响

如果需求量或各项成本在计划前的预测不准确，那么得到的订货批量就会有偏差，记偏差为δQ^*，即实际订货批量$Q'=(1+\delta)Q^*$，由总费用公式(9-1)，得

$$C(Q')=\frac{1}{2}c_1Q'+\frac{c_3D}{Q'}=\frac{1}{2}c_1(1+\delta)Q^*+\frac{c_3D}{(1+\delta)Q^*}$$

代入$Q^*=\sqrt{\dfrac{2c_3D}{c_1}}$，得

$$C(Q')=\frac{1}{2}c_1(1+\delta)\sqrt{\frac{2c_3D}{c_1}}+\frac{c_3D}{(1+\delta)}\sqrt{\frac{c_1}{2c_3D}}=\left(1+\frac{\delta^2}{2(1+\delta)}\right)\sqrt{2Dc_1c_3}=\left(1+\frac{\delta^2}{2(1+\delta)}\right)C^*$$

由于$\delta>-1$(订货量减少$\delta=-1$就是不订货)，订货量的变化会导致总成本的增加，但对总成本影响的幅度比δ小得多。

3) 各成本对订货批量和总成本的影响

设存储成本、订货成本分别增加了δ_1、δ_2，则实际存储成本、订货成本为$c_1'=(1+\delta_1)c_1$、$c_3'=(1+\delta_2)c_3$，那么实际订货批量为

$$Q'=\sqrt{\frac{2c_3'D}{c_1'}}=\sqrt{\frac{2c_3(1+\delta_1)D}{c_1(1+\delta_2)}}=\sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{1+\delta_1}{1+\delta_2}}=Q^*\sqrt{\frac{1+\delta_1}{1+\delta_2}}$$

总费用为(不包括货物购买成本)

$$C'=\sqrt{2Dc_1'c_3'}=\sqrt{2D(1+\delta_1)c_1(1+\delta_2)c_3}=\sqrt{2Dc_1c_3}\sqrt{(1+\delta_1)(1+\delta_2)}=C^*\sqrt{(1+\delta_1)(1+\delta_2)}$$

基于上述分析，以实物计量单位如件、个表示物资数量时，若Q^*不是整数，可四舍五入而取整，对总费用的影响不大。这表明经济订货批量公式是一个实用性很强的公式，在实际订货中使用非常便利。

【例9.3】 某企业全年需要钢板$D=1\,500$吨，单价5 500元/吨，存储费用c_1=50元/(年·吨)，订货费用c_3=240元/次，不允许缺货。企业管理者为改善资金使用状况，决定减少每次订货量增加订货次数，增加的总费用最好不超过5%，问每次应采购多少吨钢板？

解： 根据EOQ模型，得

订货批量$Q^*=\sqrt{\dfrac{2c_3D}{c_1}}=\left(\sqrt{\dfrac{2\times240\times1\,500}{50}}\right)$吨$=120$吨

每年的最小存储总费用$C^*=\sqrt{2c_1c_3D}=\left(\sqrt{2\times50\times240\times1\,500}\right)$元$=6\,000$元

北京大学出版社版权所有 禁止转载

若要使得增加总费用不超过 5%，令 $\frac{\delta^2}{2(1+\delta)}$=0.05，解方程得 $\delta_1 = 0.37$ 或 $\delta_2 = -0.27$，据题意应减少订货，所以取 $\delta = -0.27$，订货量为

$$Q' = (1+\delta)Q^* = (1-0.27)\times 120\text{吨} = 87.6\text{吨}$$

新的订货策略是每次订货 87.6 吨；每次钢板的购买费用为 48.18 万元；总费用为 6 300 元，比经济订货批量成本增加 300 元。

9.2.2 其他的订货批量模型

1. 模型 2：不允许缺货，生产需一定时间

1) 模型假设

在存储管理中，由于运输环节等原因的限制，经常出现的是生产需一定时间的情况，也就是说，从订购点开始的一定时间内，按照需求出库；当库存降为零时，订购的物质不断到达，一方面不断满足需求，另一方面均匀入库，入库完毕时，达到最大库存量，其余假设同模型 1。

由于这种模型最早用在确定生产批量上，故称 Production Lot Size(PLS)模型。在生产活动中，产品的生产时间是不容忽视的，即生产批量 Q 按一定的生产速度 P，需要一定的时间 t_P 方能完成，且 $P > D$；在 t_P 时间内，边以速度 P 供货(生产)，边以速度 D 消耗，t_P 时间内的进货量满足一个订货 T 的需用量，即 $Q = Pt_P = DT$，所以 $t_P = DT/P, T = Q/D$。

2) 模型建立和求解

模型的存储量变化如图 9.5 所示。

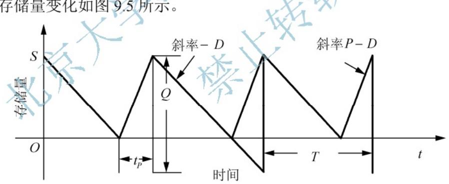


图 9.5 存储量变化过程

为了建立模型，必须首先求出平均库存量、最大库存量。由于生产批量(即订货批量) Q 需时间 t_P，单位时间的产量 P 为进货速度，故在 t_P 时间库存的实际增长速度为 $P-D$；最高库存量 $(P-D)t_P$；平均存储量 $\frac{1}{2}(P-D)t_P$；一个存储周期的存储总费用为 $\frac{1}{2}c_1(P-D)t_P T + c_3$。

将 $t_P = \frac{D}{P}T$ 代入上式，得到一个存储周期总费用为 $\frac{1}{2}c_1(P-D)\frac{D}{P}T^2 + c_3$，所以单位时间总费用为 $C_Z = \frac{1}{2}c_1(P-D)\frac{D}{P}T + \frac{c_3}{T}$。

北京大学出版社版权所有 禁止转载

将$T=\dfrac{Q}{D}$代入上式，得 $C_Z=\dfrac{1}{2}c_1(P-D)\dfrac{Q}{P}+D\dfrac{c_3}{Q}$

求导，并令 $\dfrac{\mathrm{d}C_Z}{\mathrm{d}Q}=\dfrac{1}{2}c_1\dfrac{P-D}{P}-D\dfrac{c_3}{Q^2}=0$

就可求得使单位时间总费用最低的经济订货批量为

$$Q^*=\sqrt{\frac{2c_3DP}{c_1(P-D)}}=\sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{P}{P-D}} \tag{9-4}$$

相应的订货周期为

$$T^*=\sqrt{\frac{2c_3P}{Dc_1(P-D)}}=\sqrt{\frac{2c_3}{Dc_1}}\sqrt{\frac{P}{P-D}}$$

单位时间最小存储总费用为

$$C^*=\sqrt{\frac{2Dc_1c_3(P-D)}{P}}=\sqrt{2Dc_1c_3}\sqrt{\frac{P-D}{P}} \tag{9-5}$$

最高库存量为

$$S^*=(P-D)\frac{D}{P}\sqrt{\frac{2c_3P}{Dc_1(P-D)}}=\sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{P-D}{P}}$$

【例 9.4】 某电视机厂自行生产扬声器用以装配本厂生产的电视机。该厂每天欲生产 100 台电视机，而扬声器生产车间每天可以生产 5 000 个，已知该厂每批电视机装备的生产准备费为 5 000 元，而每个扬声器一天的保管费用为 0.02 元。试确定该厂扬声器的最佳生产批量、生产时间和电视机的装配周期。

解：据题知，该模型是一个不允许缺货、边生产边装配的模型，且$D=100$，$P=5\,000$，$c_1=0.02$，$c_3=5\,000$(每次生产的准备费，电视机厂每次的订货费用)，所以最优订货批量为

$$Q^*=\sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{P}{P-D}}=\sqrt{\frac{2\times5\,000\times100}{0.02}}\sqrt{\frac{5\,000}{5\,000-100}}\approx7\,143\,(台)；$$

电视机的装配周期$T^*=\dfrac{Q^*}{D}=\dfrac{7\,143台}{100台/天}\approx71$天；

扬声器所需生产时间$t_p=\dfrac{Q^*}{P}=\dfrac{7\,143个}{5\,000个/天}\approx1.43$天。

由于$1.43\ll71$，显然扬声器达不到规模效应，客观上造成产能的浪费，这也是现代企业资源外包的原因。

2. 模型 3：允许缺货(缺货量补足)，生产时间很短

上述两个模型是以不允许缺货为前提的，但对实际的存储系统而言，由于受各种客观条件的制约，不缺货几乎是很难实现的。这样为保证不缺货，要求企业保有大量的库存，这无形中增加了存储费用开支；而缺货时，必然要求支付缺货损失费，但可以减少物资的存储量，延长订货周期，因此综合考虑存储系统的总费用，适当采取缺货策略在一定程度上是可取的。

北京大学出版社版权所有 禁止转载

1) 模型假设

一般发生缺货后的情况又可分为两种：一种是缺货后可以延期付货；另一种是发生缺货后损失无法弥补，损失顾客。由于第二种情况是企业所不希望出现的，因此在下面的讨论中，仅探讨允许延期付货的情形。在这种情况下，虽然在一段时间内发生缺货，但下批订货到达后立即补足缺货。以 t_1 表示需求全由库存现货供应的时间；允许缺货，用 t_2 表示缺货时间，且缺货部分用下一批到货一次补足；Q_1 表示缺货数量，以 c_2 表示单位缺货时间、单位数量支付的缺货损失费，如图 9.6 所示。

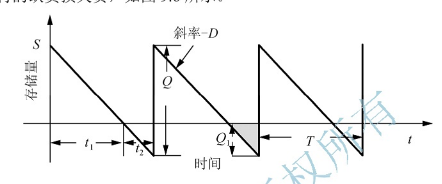


图 9.6　存储量变化过程

北京大学出版社版权所有 禁止转载

2) 模型建立和求解

从图 9.6 可以看出，由于缺货后延期交货，所以一个周期的订货批量仍然满足 $Q = DT$，$T = Q/D$；一个周期的最高存储量不是 Q，而是 $Q - Q_1$，利用三角形相似关系可以得到

$$\frac{T-t_2}{T} = \frac{Q-Q_1}{Q} \qquad \frac{t_2}{T} = \frac{Q_1}{Q} \qquad t_2 = \frac{Q_1}{Q}T$$

所以，一个存储周期内存储费用为 $\frac{1}{2}c_1(Q-Q_1)(T-t_2) = \frac{1}{2}c_1\frac{(Q-Q_1)^2}{Q}T$；订货费为 c_3；缺货费用为 $\frac{1}{2}c_2Q_1t_2$。

由于一个存储周期发生的费用总额为以上存储费、订货费及缺货费之和，则单位时间的总费用为

$$C_Z = \frac{1}{2}c_1\frac{(Q-Q_1)^2}{Q} + \frac{c_3}{T} + \frac{1}{2}\frac{c_2Q_1t_2}{T}$$

将 $T = \frac{Q}{D}$ 及 $t_2 = \frac{Q_1}{Q}T$ 代入上式，得

$$C_Z = \frac{1}{2}c_1\frac{(Q-Q_1)^2}{Q} + \frac{c_3D}{Q} + \frac{1}{2}\frac{c_2Q_1^2}{Q}$$

可见，C_Z 是一个 Q，Q_1 的二元函数，要使 C_Z 取最小，需要求 C_Z 对 Q，Q_1 的偏导数，并令其等于零，即可得经济订货批量和经济缺货量为

$$Q^* = \sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{c_1+c_2}{c_2}} \tag{9-6}$$

$$Q_1^* = \sqrt{\frac{2Dc_1c_3}{c_2(c_1+c_2)}}$$

用二阶偏导数检验可知，所求Q^*,Q_1^*使C_Z取极小，相应的订货周期为

$$T^*=\sqrt{\frac{2c_3}{Dc_1}}\sqrt{\frac{c_1+c_2}{c_2}}$$

单位时间最小存储总费用为

$$C^*=\sqrt{2Dc_1c_3}\sqrt{\frac{c_2}{c_1+c_2}} \tag{9-7}$$

最高库存量为

$$S^*=\sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{c_2}{c_1+c_2}}\text{。}$$

【例 9.5】 某工厂按照合同每月向外单位供货 100 件，每次生产准备成本为 5 元，每件年存储费为 4.8 元，每件生产成本为 20 元。若不能按期交货每件每月罚款 0.5 元，试求总费用最小的生产方案。

解：计划期为一个月，已知D=100，c_1=4.8/12=0.4，c_2=0.5，c_3=5，K=20，由上述公式得

$$Q^*=\sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{c_1+c_2}{c_2}}=\sqrt{\frac{2\times5\times100}{0.4}}\sqrt{\frac{0.4+0.5}{0.5}}\approx67\,(\text{件});$$

$$T^*=\sqrt{\frac{2c_3}{Dc_1}}\sqrt{\frac{c_1+c_2}{c_2}}=\sqrt{\frac{2\times5}{100\times0.4}}\sqrt{\frac{0.4+0.5}{0.5}}\approx0.67\,(\text{月})=21(\text{天});$$

$$C^*=\sqrt{2Dc_1c_3}\sqrt{\frac{c_2}{c_1+c_2}}=\sqrt{2\times100\times0.4\times5}\sqrt{\frac{0.5}{0.4+0.5}}\approx15\,(\text{元})\text{。}$$

北京大学出版社版权所有 禁止转载

3. 模型 4：允许缺货(需补足缺货)、生产需一定时间

1) 模型假设

假设条件除允许缺货和生产需一定时间外，其余与模型 1 相同。

2) 模型建立和求解

存储量变化如图 9.7 所示。T为一个订货周期，在t_1时间内以速度D消耗库存；在时间t_2+t_3内库存为零，Q_1是最大缺货量；在达到最大缺货量时补货到达，即以速度P进货，同时以速度D消耗供货；当缺货补足后，满足消耗后的订货进入存储，存储量以速度$P-D$增加，直至达到最大库存量，停止进货，开始下一个订货周期。

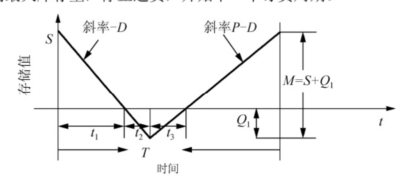


图 9.7 存储量变化过程

根据上述说明及图 9.7 可知 $T=\dfrac{M}{D}+\dfrac{M}{P-D}$ 或 $\dfrac{1}{T}=\dfrac{(P-D)D}{MP}$，又由于缺货后延期交货，所以最高存储量不是 M，而是 $M-Q_1=S$，利用三角形相似关系可以得到

$$\frac{t_2+t_3}{T}=\frac{Q_1}{M} \qquad t_2+t_3=\frac{Q_1}{M}T \qquad T-(t_2+t_3)=\frac{M-Q_1}{M}T$$

因此一个存储周期内存储费用为 $\dfrac{1}{2}c_1(M-Q_1)(T-t_2-t_3)$；订货费为 c_3；缺货费用为 $\dfrac{1}{2}c_2Q_1(t_2+t_3)$。

由于一个存储周期发生的费用总额为以上存储费、订货费及缺货费之和，那么单位时间的总费用为

$$C_Z=\frac{1}{2}c_1(M-Q_1)\frac{(T-t_2-t_3)}{T}+\frac{c_3}{T}+\frac{1}{2}\frac{c_2Q_1(t_2+t_3)}{T}$$

将 $\dfrac{1}{T}=\dfrac{(P-D)D}{MP}$、$t_2+t_3=\dfrac{Q_1}{M}T$ 及 $T-(t_2+t_3)=\dfrac{M-Q_1}{M}T$ 代入上式，得

$$C_Z=\frac{1}{2}c_1\frac{(M-Q_1)^2}{M}+\frac{c_3D(P-D)}{MP}+\frac{1}{2}\frac{c_2Q_1^2}{Q}$$

可见，C_Z 是一个 M,Q_1 的二元函数，要使 C_Z 取极小，需要求 C_Z 对 M,Q_1 的偏导数，并令其等于零，即得 M 的取值为

$$M^*=\sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{c_1+c_2}{c_2}}\sqrt{\frac{P-D}{P}}$$

注意 M^* 并不是最佳经济订货批量(图 9.7)，这是因为 M 没有考虑补货过程中的消费需求，经济订货批量 Q 仍然等于 DT，能满足一个订货周期的需求，所以

经济缺货量为

$$Q_1^*=\sqrt{\frac{2Dc_1c_3}{c_2(c_1+c_2)}}\sqrt{\frac{P-D}{P}}$$

用二阶偏导数检验可知，所求 M^*,Q_1^* 使 C_Z 取极小。

因此相应的订货周期为

$$T^*=\sqrt{\frac{2c_3}{Dc_1}}\sqrt{\frac{c_1+c_2}{c_2}}\sqrt{\frac{P}{P-D}}$$

最佳经济订货批量为

$$Q^*=DT^*=\sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{c_1+c_2}{c_2}}\sqrt{\frac{P}{P-D}} \tag{9-8}$$

单位时间最小存储总费用为

$$C^*=\sqrt{2Dc_1c_3}\sqrt{\frac{c_2}{c_1+c_2}}\sqrt{\frac{P-D}{P}} \tag{9-9}$$

最高库存量为

$$S^*=\sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{c_2}{c_1+c_2}}\sqrt{\frac{P-D}{P}}$$

北京大学出版社版权所有 禁止转载

【例 9.6】某下属车间每年能生产本厂所需某种零件 80 000 个，全厂每年需要这种零件 20 000 个。已知每个零件每月所需的存储费是 0.10 元，每批零件生产前所需的安装费用是 350 元。当供货不足时，每个零件缺货的损失费为 0.20 元/月，所缺的货到货后要补足。试问应采取怎样的存储策略最合适？并计算其最佳订货周期、最高库存量。

解：该问题属于允许缺货，生产需一定时间的模型。设计划时间期限为 1 年，因而应将有关费用的时间单位转换为年。由题知，P=80 000 个，D=20 000 个，$c_1 = 0.10\text{元/月}\times 12\text{月} = 1.2$ 元/年，$c_2 = 0.20\text{元/月}\times 12\text{月} = 2.4$ 元/年，$c_3 = 350$ 元，由上述公式得

最佳订货周期为

$$T^* = \sqrt{\frac{2c_3}{Dc_1}}\sqrt{\frac{c_1+c_2}{c_2}}\sqrt{\frac{P}{P-D}} = \sqrt{\frac{2\times 350}{20\,000\times 1.2}}\sqrt{\frac{1.2+2.4}{2.4}}\sqrt{\frac{80\,000}{80\,000-20\,000}} \approx 0.24\ (\text{年/个})$$

经济订货批量为

$$Q^* = DT^* = (20\,000\times 0.24)\text{个} = 4\,800\text{ 个}$$

最大存储量为

$$S^* = \sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{c_2}{c_1+c_2}}\sqrt{\frac{P-D}{P}} = \sqrt{\frac{2\times 350\times 20\,000}{1.2}}\sqrt{\frac{2.4}{1.2+2.4}}\sqrt{\frac{80\,000-20\,000}{80\,000}} \approx 2\,415(\text{个})$$

若产品生产速度 $P\to\infty$，缺货损失 $c_2\to\infty$ 时，$\dfrac{P-D}{P}\to 1$，$\dfrac{c_1+c_2}{c_2}\to 1$，模型 2、3、4 就变为模型 1，各个模型有关参数之间的关系见表 9-1。

北京大学出版社版权所有 禁止转载

表 9-1 模型参数表

参数 / 模型	经济订货批量 Q^*	订货周期 T^*	最小存储总费用 C^*	最大存储量 S^*
模型 1	$\sqrt{\frac{2c_3D}{c_1}}$	$\sqrt{\frac{2c_3}{Dc_1}}$	$\sqrt{2Dc_1c_3}$	$\sqrt{\frac{2c_3D}{c_1}}$
模型 2	$\sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{P}{(P-D)}}$	$\sqrt{\frac{2c_3}{Dc_1}}\sqrt{\frac{P}{(P-D)}}$	$\sqrt{2Dc_1c_3}\sqrt{\frac{P-D}{P}}$	$\sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{(P-D)}{P}}$
模型 3	$\sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{c_1+c_2}{c_2}}$	$\sqrt{\frac{2c_3}{Dc_1}}\sqrt{\frac{c_1+c_2}{c_2}}$	$\sqrt{2Dc_1c_3}\sqrt{\frac{c_2}{c_1+c_2}}$	$\sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{c_2}{c_1+c_2}}$
模型 4	$\sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{c_1+c_2}{c_2}}\sqrt{\frac{P}{P-D}}$	$\sqrt{\frac{2c_3}{Dc_1}}\sqrt{\frac{c_1+c_2}{c_2}}\sqrt{\frac{P}{P-D}}$	$\sqrt{2Dc_1c_3}\sqrt{\frac{c_2}{c_1+c_2}}\sqrt{\frac{P-D}{P}}$	$\sqrt{\frac{2c_3D}{c_1}}\sqrt{\frac{c_2}{c_1+c_2}}\sqrt{\frac{P-D}{P}}$

4. 模型 5：价格有折扣的经济批量模型

上述讨论的几个存储模型的参数都与价格没有关系，都假设价格与订货数量没有关系。然而商品有零售价、批发价和出厂价等之分，购买同一种商品的数量不同，商品的单价也不同，一般情况下购买的数量越多，商品的单价越低。下面考虑价格有折扣的存储问题。由于有价格优惠，订货时就希望多订一点货物，但多订了，存储费用必然会增加，造成资金积压。如何充分利用价格的杠杆作用，又使总费用最小，这是价格有折扣的存储模型要解决的问题。

1) 模型假设

除去货物单价随订购数量而变化外，本模型的条件皆与模型 1 的假设条件相同。

2) 模型建立和求解

记货物单价为$K(Q)$，其中Q为订货量。为便于方便，设$K(Q)$按三个数量等级变化，如图 9.8 所示。

$$K(Q)=\begin{cases}K_1, & 0\leqslant Q<Q_1\\ K_2, & Q_1\leqslant Q<Q_2\\ K_3, & Q\geqslant Q_2\end{cases}，且 K_1>K_2>K_3$$

由式(9-1)可知，单位时间内的平均总费用表示为

$$C_Z(Q)=\frac{1}{2}c_1Q+\frac{c_3D}{Q}+DK(Q)$$

若将每单位物资平均的总费用记为$C(Q)$，则

$$C(Q)=\frac{1}{2}c_1\frac{Q}{D}+\frac{c_3}{Q}+K(Q)$$

将$K(Q)$代入，得

$$C^1(Q)=\frac{1}{2}c_1\frac{Q}{D}+\frac{c_3}{Q}+K_1 \qquad Q\in[0,Q_1)$$

$$C^2(Q)=\frac{1}{2}c_1\frac{Q}{D}+\frac{c_3}{Q}+K_2 \qquad Q\in[Q_1,Q_2)$$

$$C^3(Q)=\frac{1}{2}c_1\frac{Q}{D}+\frac{c_3}{Q}+K_3 \qquad Q\in[Q_2,\infty)$$

北京大学出版社版权所有
禁止转载

如果不考虑$C^1(Q)$，$C^2(Q)$，$C^3(Q)$的定义域，它们之间只差一个常数，所以它们表示一族平行曲线，同时也表示出平均每单位货物所需要的费用$C(Q)$，如图 9.9 所示。

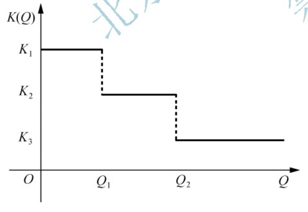


图 9.8　价格有折扣的经济批量模型

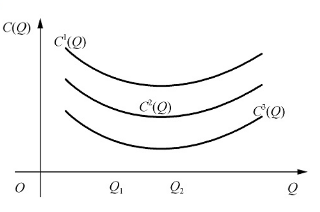


图 9.9　价格有折扣模型的成本变化示意图

为求最小总费用，先求$\dfrac{\mathrm{d}C(Q)}{\mathrm{d}Q}=\dfrac{c_1}{2D}-\dfrac{c_3}{Q^2}$，再令$\dfrac{\mathrm{d}C(Q)}{\mathrm{d}Q}=0$，得$Q_0=\sqrt{\dfrac{2c_3D}{c_1}}$，这就是模型 1 得到的经济订货批量。然而$Q_0$究竟落在哪一个区间事先难以预计，不妨设$Q_1<Q_0<Q_2$，这时也不能肯定$C^2(Q_0)$最小。从图 9.9 的直观感觉启发人们思考：是否$C^3(Q_2)$的费用更小呢？按此思路，给出价格有折扣的情况下，求解最佳订货批量Q^*的计算步骤如下。

(1) 对$C^1(Q)$(不考虑定义域)求得极值点$Q_0=\sqrt{\frac{2c_3D}{c_1}}$。

(2) $Q_0<Q_1$，则计算$C^1(Q_0)$、$C^2(Q_1)$和$C^3(Q_2)$，取其中单位货物费用最小者对应的订购批量为Q^*。

(3) 若$Q_1\leqslant Q_0<Q_2$，则计算$C^2(Q_0)$、$C^3(Q_2)$，由$\min\{C^2(Q_0),C^3(Q_2)\}$来决定$Q^*$。

(4) 若$Q_0\geqslant Q_2$，则取$Q^*=Q_0$。

这个过程首先计算每单位物资平均的总费用的共同部分$\frac{1}{2}c_1\frac{Q}{D}+\frac{c_3}{Q}$的最小值$Q_0$，若$Q_0<Q_1$，则保证了落入$Q<Q_0$、$Q_0<Q<Q_1$的订货量不是最小费用；所以经济订货批量只能等于$Q_0$或大于等于$Q_1$；而在区间$[Q_1,Q_2)$中，总费用的最小值一定在$Q_1$处达到，因为区间$[Q_1,Q_2)$在$Q_0$点的右侧，函数$\frac{1}{2}c_1\frac{Q}{D}+\frac{c_3}{Q}$是增函数，以此类推，在区间$[Q_2,\infty)$中，总费用的最小值一定在$Q_2$处达到，所以$Q_0$点的费用与各个端点处的费用相比较就能得到最小费用，从而就能确定出经济订货批量。同理，由于需求D是一个常数，函数$\frac{1}{2}c_1\frac{Q}{D}+\frac{c_3}{Q}$和$\frac{1}{2}c_1Q+\frac{c_3D}{Q}$增减区间完全一致，也可比较单位时间内的平均总费用求出经济订货批量。

【例9.7】设某车间每月需要某种零件30 000个，每次的订购费是500元，每月每件的存储费是0.2元，零件批量的单价如下：

$$K(Q)=\begin{cases}1 & 0\leqslant Q\leqslant 10\,000\\ 0.98 & 10\,000\leqslant Q\leqslant 30\,000\\ 0.94 & 30\,000\leqslant Q\leqslant 50\,000\\ 0.90 & Q\geqslant 50\,000\end{cases}$$

若不允许缺货，且一旦订货就进货，试求最佳的订货批量。

解：根据模型1，在单价不变的情况下求出最佳订购批量为

$$Q_0=\sqrt{\frac{2c_3D}{c_1}}=\sqrt{\frac{2\times 500\times 30\,000}{0.2}}\approx 12\,247\text{(个)}$$

因$10\,000\leqslant Q_0\leqslant 30\,000$，根据$C_Z(Q)=\frac{1}{2}c_1Q+\frac{c_3D}{Q}+DK(Q)$计算：

$$C_Z(12\,247)=\left(\frac{1}{2}\times 0.2\times 12\,247+\frac{500\times 30\,000}{12\,247}+0.98\times 30\,000\right)\text{元}\approx 31\,849\text{元}$$

$$C_Z(30\,000)=\left(\frac{1}{2}\times 0.2\times 30\,000+\frac{500\times 30\,000}{30\,000}+0.94\times 30\,000\right)\text{元}\approx 31\,700\text{元}$$

$$C_Z(50\,000)=\left(\frac{1}{2}\times 0.2\times 50\,000+\frac{500\times 30\,000}{50\,000}+0.90\times 30\,000\right)\text{元}\approx 32\,300\text{元}$$

比较上述单位时间内的平均总费用可知，应取Q_2=30 000为最佳订购批量。

【例9.8】某汽车制造厂年产小汽车1 000辆。每辆小汽车需配置外购发动机1台，单价为10 000元。已知发动机的经济订货量为100台/次，订货费用是3 000元/次，年保管费用率

北京大学出版社版权所有 禁止转载

为单价的 6%。供货商提出，该厂若能每次订 200 台发动机，则他们将给予优惠：发动机的单价由 10 000 元/台降至 9 500 元/台。试问该厂是否应接受此项数量折扣，将发动机的订货批量提高到 200 台/次？

解：注意该题的存储费用随着单价发生了改变，所以按照单位时间的总费用 $C_Z(Q)=\frac{1}{2}c_1Q+\frac{c_3D}{Q}+DK(Q)$ 来选择经济订货批量。

一次订货量为 100 台时：

$$C_Z(100)=\left(\frac{100\times 10\,000\times 6\%}{2}+1\,000\times\frac{3\,000}{100}+1\,000\times 10\,000\right)\text{元}=10\,060\,000\text{ 元}$$

一次订货量为 200 台时：

$$C_Z(200)=\left(\frac{200\times 9\,500\times 6\%}{2}+1\,000\times\frac{3\,000}{200}+1\,000\times 9\,500\right)\text{元}=9\,572\,000\text{ 元}$$

可知总成本最低的订货量为 200 台，故接受此项数量折扣。

9.3 单时期随机型存储模型

确定型存储模型都假设各个时期的需求是确定的，但实际问题中往往会遇到需求量不确定的情况，这是由复杂的社会现象导致的。引起需求不确定的原因很多，比如到货时间经常是不确定的，因为从订单发出，到货物送达，必定有一段时间延迟，这段延迟时间可能受生产、运输过程中许多偶然因素的影响，经常表现为一个随机变量。这就要用到随机的存储模型来解决此类存储问题，对物流企业来说，面对变化的顾客需求和不确定的物流运输环境，考察这类库存管理问题显得尤为重要。这里先探讨单时期随机型存储模型，一类需求是离散的；一类需求是连续的。

北京大学出版社版权所有 禁止转载

9.3.1 需求为随机离散的存储模型

以报童问题为例说明随机环境下的经济订货批量。设报童每天售报数量是一个随机变量，每日售出 r 份报纸的概率为 $P(r)$，根据以往的经验是已知的；每张报纸的成本为 u 元，每张报纸的售价为 v ($v>u$)元；如果报纸当天卖不出去，第二天就降低处理，设处理价为 w ($w<u$)元。问报童最好每日订购多少份报纸？

该问题就是要确定报童每天报纸的订货量 Q，目标是使赢利的期望值最大或损失的期望值最小。考虑问题的出发点：当最佳订购批量为 Q 时，无论订购量在增加一个或减少一个，都会使得费用损失(缺货费和订购过量损失之和)增加。

售出报纸数量为 r 的概率为 $P(r)$，$\sum_{r=1}^{\infty}P(r)=1$，设报童订购报纸数量为 Q，这时的损失有以下两种。

(1) 当供大于求($Q\geqslant r$)时，报纸因当天不能售完，第二天需降价处理，每件降价损失为 $u-w$，其损失的期望值为

$$\sum_{r=0}^{Q}(u-w)(Q-r)P(r)$$

(2) 当供不应求($Q < r$)时，因缺货而失去销售机会，每件缺货损失为$v-u$，其损失的期望值为

$$\sum_{r=Q+1}^{\infty}(v-u)(r-Q)P(r)$$

所以损失的期望值为

$$\begin{aligned}C(Q) &= \sum_{r=0}^{Q}(u-w)(Q-r)P(r) + \sum_{r=Q+1}^{\infty}(v-u)(r-Q)P(r) \\ &= (u-w)\sum_{r=0}^{Q}(Q-r)P(r) + (v-u)\sum_{r=Q+1}^{\infty}(r-Q)P(r)\end{aligned} \tag{9-10}$$

由于报童订购报纸的份数Q只能取整数，需求量r也只能取整数，不能用微积分的方法求式(9-10)的极值。若设报童每天订购的报纸的最佳批量为Q^*，无论订购量再增加一个或减少一个，都会使得费用损失(缺货费和订购过量损失之和)增加。

则必有

$$\begin{cases}C(Q^*) \leqslant C(Q^*+1) \\ C(Q^*) \leqslant C(Q^*-1)\end{cases}$$

同时成立，故可将上述两不等式联立求解可得最佳批量Q^*。

考虑$C(Q) \leqslant C(Q+1)$应满足的条件为

$$\begin{aligned}&(u-w)\sum_{r=0}^{Q}(Q-r)P(r) + (v-u)\sum_{r=Q+1}^{\infty}(r-Q)P(r) \\ &\leqslant (u-w)\sum_{r=0}^{Q+1}(Q+1-r)P(r) + (v-u)\sum_{r=Q+2}^{\infty}(r-Q-1)P(r)\end{aligned}$$

化简后，得以下表达式

$$(v-w)\sum_{r=0}^{Q}P(r) - (v-u) \geqslant 0$$

即

$$\sum_{r=0}^{Q}P(r) \geqslant \frac{v-u}{v-w} \tag{9-11}$$

考虑$C(Q) \leqslant C(Q-1)$应满足的条件为

$$\begin{aligned}&(u-w)\sum_{r=0}^{Q}(Q-r)P(r) + (v-u)\sum_{r=Q+1}^{\infty}(r-Q)P(r) \\ &\leqslant (u-w)\sum_{r=0}^{Q-1}(Q-1-r)P(r) + (v-u)\sum_{r=Q}^{\infty}(r-Q+1)P(r)\end{aligned}$$

化简后，得

$$(v-w)\sum_{r=0}^{Q-1}P(r) - (v-u) \leqslant 0$$

即

$$\sum_{r=0}^{Q-1}P(r) \leqslant \frac{v-u}{v-w} \tag{9-12}$$

北京大学出版社版权所有
禁止转载

所以报童应该准备的报纸最佳数量 Q 应该是按下述不等式来确定：

$$\sum_{r=0}^{Q-1} P(r) < \frac{v-u}{v-w} \leqslant \sum_{r=0}^{Q} P(r) \tag{9-13}$$

从而可由式(9-13)确定最佳订购批量 Q^*，其中 $\frac{v-u}{v-w}$ 称为临界值，$v-u$ 是每件缺货损失费用，$v-w=(v-u)+(u-w)$ 是每件缺货损失和降价损失费用之和。

【例 9.9】 设某货物的需求量为 19～28 件，已知需求量 r 的概率分布见表 9-2，并知其成本为每件 5 元，售价为每件 10 元，处理价为每件 2 元。问应进货多少，能使总利润的期望值最大？

表 9-2　需求量的概率分布

需求量 r	19	20	21	22	23	24	25	26	27	28
概率 $P(r)$	0.12	0.18	0.23	0.13	0.10	0.08	0.05	0.04	0.04	0.03

解： 该题属于单时期需求离散随机变量的存储模型。据题知 $u=5$，$v=10$，$w=2$，由式(9-13)得

$$\sum_{r=19}^{Q-1} P(r) \leqslant \frac{10-5}{10-2} \leqslant \sum_{r=19}^{Q} P(r)$$

即

$$\sum_{r=19}^{Q-1} P(r) \leqslant 0.625 \leqslant \sum_{r=19}^{Q} P(r)$$

因为 $P(19)=0.12$，$P(20)=0.18$，$P(21)=0.23$，$P(22)=0.13$，而且

$$P(19)+P(20)+P(21)=0.53<0.625$$

$$P(19)+P(20)+P(21)+P(22)=0.66>0.625$$

所以最佳订货批量 $Q^*=22$ 件。

【例 9.10】 某商场拟在新年期间出售一批日历画片，每出售 1 000 张可以赢利 70 元，如果在新年期间不能售出，必须削价处理(削价一定可以售出)，此时每张损失 0.040 元。根据以往的经验，市场需求的概率见表 9-3，问年前一次订货应订购画片几千张？

表 9-3　需求量的概率分布

需求数量 r/千张	0	1	2	3	4	5
需求概率 $P(r)$	0.05	0.10	0.25	0.35	0.15	0.10

解： 该题属于单时期需求离散随机变量的存储模型。据题知 $v-u=70$，$u-w=40$，由式(9-13)，得

$$\frac{v-u}{v-w}=\frac{v-u}{v-u+u-w}=\frac{70}{70+40}=0.64$$

由于 $\sum_{r=0}^{2} P(r)=0.40$，$\sum_{r=0}^{3} P(r)=0.75$，因此 $Q^*=3\,000$，即应该订购 3 000 张日历画片。

这类模型还可以考虑日历、杂志、各类季节性的货物、时装等商品。它们的共同点是：如果一次进货过多，就会有多余商品销不出去，多余物品要贬值处理；如果进货不足，发

北京大学出版社版权所有
禁止转载

生脱销，就会失去销售时机。

9.3.2 需求为随机连续的存储模型

设有某种单时期需求的物资，需求量r为连续随机变量，已知其密度函数为$\varphi(r)$，则$\varphi(r)\mathrm{d}r$表示随机变量在r与$r+\mathrm{d}r$之间的概率，其分布函数$F(a)=\int_0^a \varphi(r)\mathrm{d}r$。设每件物品的成本为$u$元，售价为$v$元($u<v$)，如果当期销售不出去，下一期就要降价处理，设处理价为w元($w<u$)，试求最佳订货批量Q^*。

同需求为离散随机变量一样，如果订货量大于需求量($Q\geqslant r$)时，其赢利的期望值为$\int_0^Q [(v-u)r-(u-w)(Q-r)]\varphi(r)\mathrm{d}(r)$；当订货量小于需求量($Q<r$)时，其实际的销售量为$\min\{r,Q\}=Q$，赢利的期望值为$\int_Q^\infty (v-u)QP(r)$，故总利润的期望值为

$$
\begin{aligned}
\mathrm{E}(Q)&=\int_0^Q [(v-u)r-(u-w)(Q-r)]\varphi(r)\mathrm{d}(r)+\int_Q^\infty (v-u)QP(r)\\
&=-uQ+(v-w)\int_0^Q r\varphi(r)\mathrm{d}(r)+w\int_0^Q Q\varphi(r)\mathrm{d}(r)\\
&\quad+v\left[\int_0^\infty Q\varphi(r)\mathrm{d}(r)-\int_0^Q Q\varphi(r)\mathrm{d}(r)\right]\\
&=(v-u)Q+(v-w)\int_0^Q r\varphi(r)\mathrm{d}(r)-(v-w)\int_0^Q Q\varphi(r)\mathrm{d}(r)
\end{aligned}
$$

利用含参变量积分求导公式为

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_a^{b(t)} f(x,t)\mathrm{d}x=\int_a^b f_t'(x,t)\mathrm{d}x+f(b,t)\frac{\mathrm{d}b(t)}{\mathrm{d}t}$$

得

$$
\begin{aligned}
\frac{\mathrm{dE}(Q)}{\mathrm{d}Q}&=(v-u)+(v-w)Q\varphi(Q)-(v-w)\left[\int_0^Q \varphi(r)\mathrm{d}r+Q\varphi(Q)\right]\\
&=(v-u)+(v-w)\int_0^Q \varphi(r)\mathrm{d}r
\end{aligned}
$$

令$\frac{\mathrm{dE}(Q)}{\mathrm{d}Q}=0$，得

$$\int_0^Q \varphi(r)\mathrm{d}r=\frac{v-u}{v-w}$$

记$F(Q)=\int_0^Q \varphi(r)\mathrm{d}r$，则有

$$F(Q)=\frac{v-u}{v-w} \tag{9-14}$$

又因

$$\frac{\mathrm{d}^2\mathrm{E}(Q)}{\mathrm{d}Q^2}=-(v-w)\varphi(Q)<0$$

故式(9-14)式求出的Q^*为$\mathrm{E}(Q)$的最大值点，即Q^*是使总利润的期望值最大的经济批量，同时可以看出式(9-14)与式(9-13)式是一致的。

【例 9.11】 某时装商店计划冬季到来之前订购一批款式新颖的皮制服装。每套皮装进价是1 000 元，估计可以获得 80%的利润，冬季一过则只能按进价的 50%处理。根据市场需求

北京大学出版社版权所有 禁止转载

预测，该皮装的销售量服从参数为1/60的指数分布，求该时装店的最佳订货批量。

解：该题属于单时期需求是连续随机变量的存储模型。据题知

$$v-u=1\,000\times 80\%=800\text{，}\quad u-w=1\,000\times 50\%=500$$

根据式(9-14)，$F(Q)=\dfrac{800}{800+500}=0.615\,4$。

由于皮装的销售量服从参数为1/60的指数分布，则其分布密度函数为

$$\varphi(x)=\begin{cases}\dfrac{1}{60}\mathrm{e}^{-\frac{1}{60}x}, & x>0\\ 0, & x\leqslant 0\end{cases}$$

由于 $\int_0^Q \dfrac{1}{60}\mathrm{e}^{-\frac{1}{60}x}\mathrm{d}x=1-\mathrm{e}^{-\frac{Q}{60}}=0.6154$，因此 $Q^*=-60\times\ln(1-0.615\,4)\approx 57$(件)。

【例 9.12】 某大学内一书刊亭经营某种期刊，每册进价 0.8 元，售价 1.00 元，如过期，处理价为 0.5 元。据多年实践经验统计，需求服从均匀分布，最高需求量 $b=1\,000$ 册，最低需求量 $a=500$ 册，问应进货多少，才能保证期望利润最高？

解：该题属于单时期需求是连续随机变量的存储模型。据题知

$$v=1.0\text{，}\quad u=0.8\text{，}\quad w=0.5$$

根据式(9-14)，$F(Q)=\dfrac{v-u}{v-w}=\dfrac{1.0-0.8}{1.0-0.5}=0.40$。

由概率论可知，均匀分布的密度函数为

$$\varphi(x)=\begin{cases}\dfrac{1}{b-a} & a\leqslant x\leqslant b\\ 0 & \text{其他}\end{cases}$$

令 $\int_0^Q\varphi(x)\mathrm{d}x=\int_a^Q\dfrac{1}{b-a}\mathrm{d}x=\int_{500}^Q\dfrac{1}{1\,000-500}\mathrm{d}x=\dfrac{Q-500}{500}=0.40$，得此书刊亭的最佳订货批量为 $Q^*=700$ 册。

北京大学出版社版权所有 禁止转载

9.4 多时期随机型存储模型

多时期存储模型是最接近于现实环境的存储系统，这是一类考虑时间因素的随机动态库存模型，它与单时期库存模型的不同之处在于每个周期的期末库存货物对于下周期仍然可用。这里简要介绍存储系统的两个基本概念和存储策略的应用。

9.4.1 再订货点和安全库存

前面讲述的确定型存储模型都做了一些假定，有的假设需求量保持不变；有的假设订货后在很短时间内按时交货。这些假设都是非常苛刻的，因为物流系统的复杂性，在实际的管理运作中，这些假设几乎都不能满足，如假设很短时间内到货，实际中往往需要一定的生产或运输时间，若这些时间能够清晰地确定下来，就可以保证存储系统的库存基本按规定日期得到补充，这就是存储系统中订货点的提前问题。

然而，物流系统常常受各种因素的影响，订货不能按时送达，发生随机的延迟拖后，从而发生缺货现象。确定型存储模型不考虑这种随机因素，不能描述这些随机因素对订货

批量的影响，只有用随机型存储模型才能分析上述现象。一般为了消除或弥补这种随机波动的影响，需要对需求量和提前期的历史资料进行统计分析，在存储模型订货批量的基础上添加一个安全库存。

1. 再订货点

若记订货到收货之间的时间为t，为保证这段时间需求消耗，那么就不能等到存量为零时再去订货，否则就会发生缺货。为了保证这段时间的存量不小于零，必须提前进货，这段时间也叫订货提前期，当存量降到什么水平就要提出订货，这一水平就称为再订货点。

对于确定型存储模型来说，再订货点$s = tD$，即当库存降到s时发出申请订货的信号。

【例 9.13】某企业全年需要某种材料 1 000 吨，单价为 500 元/吨，每吨保管费用为 50 元，每次订货手续费为 170 元，如果订货提前期为 10 天，求再订货点s。

解：据题知$D = 1\,000$吨，$t = 10$天=0.027 年，则$s = tD = 1\,000\text{吨/年} \times 0.027\text{年} = 27$吨，即库存降到 27 吨时开始订货。

若通过连续盘存法获得库存降到 27 吨，采购部门就开始订货。

2. 安全库存

安全库存量又称为保险库存量，是一项以备应付缺货情况的库存量，它是一种额外的库存量，在正常情况下不必动用它，只有出现缺货时才动用它。安全库存的使用是为了增加系统对环境的适应性和供应的稳定性，比如在一个生产存储系统中，由于生产不均衡，需求的各种变化，使存货也发生变化。在订货方式中，每当库存量减至s时，就按一定批量订货补充。如果订货后立即交货，如图 9.10 中A，并不动用安全库存量；如果订货后不能按时交货，出现延误时间，如图 9.10 中的C，将要动用安全库存，以应付延迟时间内的用量；如果在订货到交货期间，出现过量使用，安全库存量消耗殆尽，如图 9.10 中的B，也需动用安全库存。

北京大学出版社版权所有 禁止转载

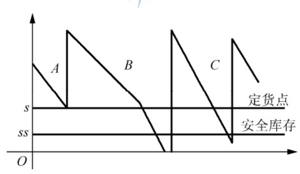


图 9.10　存储量变化过程

安全库存在正常情况下一般不用，一经动用，则应在下批订货到达时立即补齐。令订货提前期为t，由于单位时间的需求量为D，则在订货提前期内的需求量为Dt。若t和D服从某种统计分布，有关它们的统计资料比较可靠和完备，则可运用数理统计的有关方法，从满足预订的某一不缺货概率(服务水平)出发，来确定必要的安全库存ss：

$$ss = \beta\sigma_{Dt} = \beta\sqrt{\bar{t}\sigma_D^2 + \bar{D}^2\sigma_t^2} \tag{9-15}$$

式中　β——安全系数，它可根据预定的不缺货概率(服务水平$1-\alpha$)求出，其值满足概率

等式 $P(X>\beta)=\alpha$；$\bar{t},\sigma_t^2$——订货提前期 t 的均值和方差；$\overline{D},\sigma_D^2$——单位需求量 D 的均值和方差。

当供货条件较为稳定，订货提前时间基本上确定不变，而需求量是随机变化的，这时 $\sigma_t^2=0$，$\sigma_{Dt}^2=\bar{t}\sigma_D^2$，安全库存简化为

$$ss=\beta\sigma_{Dt}=\beta\sqrt{\bar{t}\sigma_D^2}=\beta\sigma_D\sqrt{\bar{t}}$$

当提前期内的顾客需求情况是确定的常数，而提前期是随机变化的，这时 $\sigma_D^2=0$，安全库存简化为

$$ss=\beta\sigma_{Dt}=\beta\overline{D}\sigma_t$$

实践表明，很多物资订购期间实际需求量出现的概率服从正态分布，可以应用正态分布的有关理论来确定安全系数 β，表 9-4 给出了正态分布的安全系数，允许缺货概率可根据企业长期经营的经验做概略规定，如服务水平不低于 98%，即表示在 100 个订货期内，允许缺货次数不得多于 2 次。

北京大学出版社版权所有

表 9-4　安全系数表

服务水平	0.99	0.98	0.95	0.90	0.80	0.70
安全系数	2.33	2.05	1.65	1.29	0.84	0.53

【例 9.14】 某物资仓库对过去 50 个订货提前期的实际需求量 $(Dt)_i$ 进行的统计分析，见表 9-5，如要求服务水平不低于 98%，试确定安全库存量。

禁止转载

表 9-5　需求量分布

实际需求量 $(Dt)_i$	70	80	90	100	110	120	130
频数 f_i	1	2	9	25	10	2	1

解：不妨设原始数据服从正态分布，计算

(1) 订货期间实际需求量的期望和标准差为

$$\overline{Dt}=\frac{70\times1+80\times2+\cdots+130\times1}{1+2+9+25+10+2+1}=100$$

$$\sigma_{Dt}=\sqrt{\frac{(70-100)^2\times1+(80-100)^2\times2+\cdots+(130-100)^2\times1}{1+2+9+25+10+2+1}}=\sqrt{\frac{5\,300}{50}}=10.3\,(件)$$

(2) 根据服务水平不低于 98%，查正态分布表，得 $\beta=2.05$，安全库存量：

$$ss=\beta\sigma_{Dt}=2.05\times10.3件=21件$$

从表 9-5 可以看出，如安全库存量为 21 件时，只有当订货期间的实际需求出现 130 件时才会发生缺货，其缺货概率为 1÷50＝2%，不缺货概率为 98%，因此安全库存量为 21 件时，就可以满足服务水平不低于 98%的要求。

【例 9.15】 某商店的可乐日平均需求量为 10 箱，顾客的需求服从标准差为 2 箱/天的正态分布，提前期满足均值为 6 天、标准差为 1.5 天的正态分布，并且日需求量与提前期是相互独立的，试确定 95%的顾客满意度下的安全库存量。

解：由题意得 $\overline{D}$=10 箱/天，$\sigma_D=2$ 箱/天，订货提前期的 $\bar{t}$=6，σ_t=1.5 天，服务水平为 0.95 对应的 β=1.65，代入公式(9-15)，得

$$ss = \beta\sqrt{\bar{t}\sigma_D^2 + \overline{D}^2\sigma_t^2} = \left(1.65 \times \sqrt{6 \times 2^2 + 10^2 \times 1.5^2}\right)箱 = 26箱$$

即在满足95%的顾客满意度的情况下，安全库存是26箱。

9.4.2 存储系统的存储策略

对于多时期存储系统来说，要解决的仍然是两个基本问题，即何时订货及其每次订多少货的问题。在实际应用中，存储管理人员根据不同物质的需求特点，本着经济费用原则或存储系统对服务水平的要求，对不同的物质采取不同的存储策略。这里简要介绍几种不同的存储策略，以窥存储系统管理之全貌。

1. (s, S) 存储策略

对 (s, S) 存储策略来说，s 是存储量下限，每阶段初检查库存量，当库存 $I \leqslant s$ 时，需要订货；当库存 $I > s$ 时，本阶段不订货。S 是存储量的上限，每次订货时，使库存量达到存储量 S，每次订货批量为 $Q = S - I$。

若假设订货后立刻就能送到，即订货提前期为零，所以不需设立安全库存，这时多时期与单时期随机型存储模型的假设一致，系统的最大库存量可同理求得。

【例 9.16】 设某企业对于某种材料每月需求量的资料见表9-6。每次订货费为400元，每月每吨保管费为40元，每月每吨缺货费用为1 400元，每吨材料的购置费为752元，该企业欲采用 (s,S) 库存策略来控制库存量，试求出 S 之值；若订货时的库存 I_0=40，则订货批量为多少？

表9-6 需求量概率分布表

需求量 x_i/吨	55	64	75	82	88	90	100	110
概率 $P(x_i)$	0.05	0.10	0.15	0.15	0.20	0.10	0.15	0.10
累积概率	0.05	0.15	0.30	0.45	0.65	0.75	0.90	1.00

解：由题知 $c_3 = 400$ 元，$K = 752$ 元，$c_1 = 40$ 元，$c_2 = 1\,400$ 元

临界值 $\dfrac{c_2 - K}{c_2 + c_3} = \dfrac{1\,400 - 752}{1\,400 + 40} = 0.45$

由 $\sum_{i=1}^{3} P(x_i) < 0.45 \leqslant \sum_{i=1}^{4} P(x_i)$，得最大订货量

$$S = x_4 = 82\text{ 吨}$$

因为 $I_0 = 40$ 吨，则订货量 $Q = S - I_0$=82−40=42，即需补充42吨货物。

【例 9.17】 某商场经销一种电子产品，根据历史资料，该产品的销售量服从在区间[50,100]的均匀分布，每台产品进货价为3 000元，单位库存费为40元，若缺货商店为了维护自己的信誉，将以每台3 400元向其他商店进货后再卖给顾客，每次订购费用为400元，设期初无库存，试确定最佳订货量及 S 值。

解：由题知 $K = 3\,000$ 元/台，$c_1 = 40$ 元/(台·年)，$c_2 = 3\,400$ 元/台，$c_3 = 400$ 元/次，$I_0 = 0$

临界值 $\dfrac{c_2 - K}{c_1 + c_2} = \dfrac{3\,400 - 3\,000}{3400 + 40} = 0.116\,3$

北京大学出版社版权所有 禁止转载

又因为销售量服从均匀分布，所以其密度函数为 $\varphi(x)=\begin{cases}\dfrac{1}{50} & 50\leqslant x\leqslant 100\\ 0 & 其他\end{cases}$

由 $\int_0^S \varphi(x)\mathrm{d}x=\int_{50}^S \frac{1}{50}\mathrm{d}x=\frac{S-50}{50}=0.1163$，得

$$S\approx 56\text{台},\qquad Q=S-I_0=56\text{台}$$

2. (Q,s) 存储策略

实际中也存在供货市场和需求市场都是随机的存储系统，一般称为双随机存储系统。双随机的库存系统的求解一般是较为困难的，但实际工作中还是常用计算机模拟这种复杂的库存系统的控制或采用一些较简单的带有某些经验性质的方法。

若随机型存储系统的需求量和订货提前期都是随机的，设 D 为单位需求量，其他有关参数与前面介绍模型一致，则推导过程与确定性存储模型同样，可得经济订货批量公式为

$$Q^*=\sqrt{\frac{2c_3D}{c_1}}$$

其中 c_3 为一次定购费，c_1 为该种物资单位时间的存储费用。

为在某服务水平下提供安全保证，可将安全库存量加到正常存货中以提供所希望达到的服务水平(即不缺货的概率)。这时有 $s=l_{Dt}+\beta\sigma_{Dt}$，式中 s 为订货点，l_{Dt} 和 σ_{Dt} 分别为订货期内的销售量 Dt 的均值与标准差，β 为安全库存系数，$\beta\sigma_{Dt}$ 为安全库存量。因此订货策略：当订货提前期大于零时，若存储量降低到 s，则以 Q^* 为订货量进行订货。

【例 9.18】 设某公司订购一种备件，采取 (Q,s) 存储策略。一次订货费为 60 元，年平均需求量为 500 件，每件年存储费为 40 元，提前期的销售量服从正态分布 $N(15,2^2)$。为使不缺货的概率达到 99%，问总费用最小的 Q 和 s 各为多少？

解：据题知 D=500 件/年，c_1=40 元/(件·年)，c_3=60 元，则

$$Q^*=\sqrt{\frac{2c_3D}{c_1}}=\left(\sqrt{\frac{2\times 60\times 500}{40}}\right)\text{件}\approx 39\text{件}$$

由于不缺货的概率达到 99%，查表 9-4 得 $\beta=2.33$，订货点为

$$s=l_{Dt}+\beta\sigma_{Dt}=(15+2.33\times 2)\text{件}\approx 20\text{件}$$

故再订货点为 20 件，每次订货 39 件。

3. (T,S) 存储策略

前面探讨的两种策略属于定量存储策略，这里分析定期存储策略 (T,S)，即按照一定固定时间间隔的日期办理订货，补充库存。

由于需求量和订货提前期都是随机的，因此 (T,S) 存储策略的每次订货的实际订货量不是常量，要根据最高库存量 S 和在库库存量 I，以及未到的在途库存量 Q_w 及其发生缺货的数量 Q_s 来确定。

1) 订货间隔期 T

(T,S) 存储策略的两次订货时间间隔固定，由存储系统的参数确定，可按订货批量 Q 和订货周期确定订货间隔 T，即

北京大学出版社版权所有 禁止转载

$$Q=\sqrt{\frac{2c_3\overline{D}}{c_1}} \qquad T=\frac{Q}{\overline{D}}$$

当订货周期需要和供货厂商共同商定时，可以根据订货时间间隔T协调工厂的生产计划，合理确定订货间隔；也可根据物质供应情况统计资料，求得平均订货周期。

2) 最大库存量S

最大库存量S需满足订货间隔期T和订货提前期t的需求，以及保证一定的服务水平。这样最大库存量由两部分组成，一是$T+t$时间内的需求，一是$T+t$时间内安全库存。

(T, S)存储策略的安全库存为

$$ss=\beta\sigma_{D(T+t)}=\beta\sqrt{\overline{D}^2\sigma_{T+t}^2+(T+\bar{t})\sigma_D^2}$$

因为T一旦确定，就是一个常量，所以$\sigma_{T+t}^2=\sigma_t^2$，这样

$$ss=\beta\sqrt{\overline{D}^2\sigma_t^2+(T+\bar{t})\sigma_D^2}$$

若供货条件较为稳定，订货时间基本为一个常数时，上式可以简化为

$$ss=\beta\sqrt{(T+\bar{t})}\cdot\sigma_D$$

则最高库存量$S=\overline{D}(T+\bar{t})+ss=\overline{D}(T+\bar{t})+\beta\sqrt{\overline{D}^2\sigma_{T+t}^2+(T+\bar{t})\sigma_D^2}$。

若在订货时，在库库存为I，已订货尚未到达的库存为Q_w，则实际订货量为

$$Q^*=S-I-Q_w$$

【例 9.19】 设某公司对一种零件的月平均需求量为 248 件，标准差 80 件；订货提前期t的均值为 1 月，标准差σ_t=1.2 月。已知每次订货费为 112 元，月存储费为 0.3 元/个。现库存管理采用(T, S)策略，安全系数$\beta=1.65$，求订货周期和最高库存量。

解：先确定订货周期T。

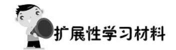
应用经济批量公式，得

$$Q^*=\sqrt{\frac{2c_3\overline{D}}{c_1}}=\sqrt{\frac{2\times112\times248}{0.3}}\approx430\,(\text{件})$$

$$T=\frac{Q^*}{\overline{D}}=\left(\frac{430}{248}\right)\text{月}\approx1.74\text{ 月}$$

最高库存量为

$$\begin{aligned}S&=\overline{D}(T+\bar{t})+ss=\overline{D}(T+\bar{t})+\beta\sqrt{\overline{D}^2\sigma_{T+t}^2+(T+\bar{t})\sigma_D^2}\\&=\left[248\times(1.74+1)+1.65\times\sqrt{248^2\times1.2^2+2.74\times6\,400}\right]\text{件}\\&\approx1\,217\text{件}\end{aligned}$$



VMI 库存管理方式

1958 年，Magee 首次提出供应商管理库 VMI(Vendor Managed Inventory)概念，它是以用户和供应商双方获得最低成本为目的，在一个共同的协议下由供应商管理库存，并不断监督协议的执行情况和修改

北京大学出版社版权所有 禁止转载

协议内容，使库存管理得到持续改进的合作性策略。供应商收集分销中心、仓库和 POS 数据，实现需求和供应相结合，下游企业只需要帮助供应商制订计划，从而下游企业实现零库存，供应商的库存也大幅度减少。

VMI 是一种很好的库存管理策略，它能够突破传统的条块分割的管理模式，以系统的、集成的管理思想进行库存管理，使供应链系统能够获得同步化的运作。VMI 改变了过去由下游用户进行采购决策的供应模式，而是由卖方直接决定对下游用户的配货，且卖方会给下游用户一个不缺货的承诺，其交换条件是下游用户必须随时为卖方提供终端信息。VMI 使下游用户省去了很多麻烦(频繁监督库存、订货，尤其是对那些库存品种有成千上万的零售商而言)及库存风险和损失，使卖方知道其产品的终端销售信息(跨过下游用户)，利于准确地预测、计划，实现了 JIT，可以减少“牛鞭效应”。

本 章 小 结

存储活动是整个物流活动的关键环节之一。本章从存储系统的结构切入，首先介绍了存储系统的运营费用、存储策略和模型分类等有关内容；接着对各种确定型存储模型进行了费用分析，得出了著名的订货批量公式；然后就需求是离散或连续的两类单时期随机型存储模型展开了深入的探讨；最后结合多时期随机型存储模型，给出了再订货点、安全库存等概念，并就一些具体的存储策略在存储系统中的应用进行了描述。

北京大学出版社版权所有 禁止转载



关键术语(中英文)

存储论(Inventory Theory)
经济订货批量(Economic Ordering Quantity)
存储策略(Inventory Police)
确定存储模型(Deterministic Inventory Model)
订货点(Ordering Point)
存储系统(Inventory System)
参数分析(Parametric Analysis)
生产批量模型(Production Lot Size)
随机存储模型(Stochastic Inventory Model)
报童问题(Newsboy Problem)



知识链接

存储论——肯尼思·约瑟夫·阿罗(Kenneth Joseph. Arrow)

肯尼思·约瑟夫·阿罗 1921 出生于美国纽约，1936 年考上了纽约市社会科学学院，经过四年的学习，他又考进了哥伦比亚大学继续深造，仅用一年的时间就获得了哥伦比亚大学文科硕士学位。在学习中他认识到：研究现代经济必须有深厚的数学基础。因此，阿罗一直没有放松学习数学。他攻读了微积分、线性代数等高等数学课程。



1942 年，获得了硕士学位的阿罗继续攻读博士学位。可是，第二次世界大战爆发了，美国政府开始大量征兵。年满 21 岁的阿罗被应征入伍当了空军。在服兵役的四年中，他始终没有放弃学习，只要有空闲，他就看书学习，研究卡尔多、希克斯、伯格森等许多经济学家的著作。1946 年，阿罗退伍后，又重新投身于经济学的研究中。

阿罗从事的专业是经济学和运筹学。他研究的经济理论是社会选择论、一般均衡论、资源配置的静态与动态理论、不稳定性经济学等，特别着重研究个人决策、信息和组织。他利用数学分析工具，研究一般均衡和福利经济学，对资产阶级经济学说新的发展作出了重要贡献。阿罗是美国经济学家，斯坦福大学营运研究名誉教授，于1972年因在一般均衡理论方面的突出贡献与约翰·希克斯共同荣获诺贝尔经济学奖。

阿罗的主要著作有：《社会选择与个人价值》《线性规划与非线性规划研究》《存货与生产的数学理论》《对企业需求的时间体系分析》《公共投资、收益率和最适度财政政策》《一般竞争分析》《风险负担理论文集》《组织的限制》《资源分配过程研究》。此外，还有《最优投资政策》等约100多篇论文，载于各种学术刊物。



习题9

9.1 填空题

(1) 存储系统的输出的方式可能是均匀连续的，也可能是__________的。

(2) 设某工业企业年需钢材1 200吨，分三次订货，则平均库存量为__________。

(3) 在确定性存储问题中，设订货费为c_3，单位存储费为c_1，缺货费为c_2，需求率为D，生产速度率为$P\,(P>D)$，且无提前期。则在不允许缺货时，最佳的经济订货批量为$Q^*=$__________；允许缺货时最佳订货批量为$Q^*=$__________。

(4) 存储成本和订货成本同时增加i%，则总成本也增加__________。

(5) 在报童问题中，若卖不完的报纸退回报社的价格由0.2元降至0.1元，问在其他条件均不变的情况下报纸的准备量应该__________。

9.2 判断题

(1) 一般来说，一个存储系统的费用主要包括存储费和订货费。 ( )

(2) (s,Q)策略是连续盘存策略，当存储量降到s时立即提出订货，订货量等于$s+Q$。 ( )

(3) 在不允许缺货的情况下，在费用上处理的方式是将缺货损失费视为无穷小。( )

(4) 在确定型的4种模型中，允许缺货、边供应边需求订货策略的总成本最低。( )

(5) 在遇到随机性存储问题时，企业选取的存储策略的优劣，通常是以赢利的期望值的大小或损失期望值的大小作为衡量的标准。 ( )

9.3 某机械部件每件进厂价为500元，年需求总额为60万元，求得最佳订货批量为300件，年保管费用率为12%。求按经济订货批量进货时，年订货多少次，每次订货费用、年保管费用和年总存货费用各是多少？

9.4 某印刷厂负责印刷一本年销售量为120万册的书，该厂每天的生产能力是几十万册，该书的销售是均匀的。若该厂只按每天销售印刷，则可使生产率与销售率同步，从而无库存，但每天印完此书又得换印刷别的书，其生产调节费为每天2 000元。每万册书储存一天的费用为4.53元，缺货一天的损失为1.02元，试分析比较缺货与不缺货的最优策略哪个比较好，并说明理由。

9.5 设需要某物品12件/天，不允许缺货，存储费率为0.02元/(件·天)。为满足需要，可以采取订购或者自行生产。有关数据见表9-7。试决定经济的物品供应来源：订购或自行生产？经济批量与存储水平各为多少？

北京大学出版社版权所有 禁止转载

表 9-7 有关数据

	订　购	自行生产
提前期或生产准备期/天	8	13
物品单价/(元/件)	11	9.6
每次订购费或准备费/元	20	90
补充速率/(件/天)	∞	25

9.6 某厂下一年度需用某种建材 1 000 吨，单价是 1 000 元/吨。已知经济订货量为 100 吨/次，订货费用是 500 元/次，年保管费用率为 10%。建材供货商提出，该厂若能每次订 200 吨，则他们将给予优惠：单价由 1 000 元/吨降至 900 元/吨。假定不计建材保管损耗，试问该厂是否应接受此项数量折扣，将建材的订货批量提高到 200 吨/次？

9.7 某货物的需求量在 14～21 件，其概率分布见表 9-8，每卖出一件可赢利 6 元，每积压一件，损失 2 元，问一次性进货多少件，才使赢利期望最大？

表 9-8 某货物的概率分布

需求量	14	15	16	17	18	19	20	21
概率	0.10	0.15	0.12	0.12	0.16	0.18	0.10	0.07
累积概率	0.10	0.25	0.37	0.49	0.65	0.83	0.93	1.00

北京大学出版社版权所有 禁止复制

9.8 某设备上有一关键零件常需更换，更换需要量 x 服从泊松分布，根据以往的经验平均需要量为 5 件，此零件的价格为 100 元/件，若零件用不完，到期末就完全报废，若备件不足，待零件损坏了再去订购就会造成停工损失 180 元，问应备多少备件最好？

9.9 某公司利用塑料制成产品出售，已知每箱塑料购价为 800 元，订购费为每次 60 元，存储费为每箱 40 元，缺货费为每箱 1 015 元，原有存储量 10 箱，已知对原料需求的概率是 $P(30)=0.20$，$P(40)=0.20$，$P(50)=0.40$，$P(60)=0.20$，求该公司的(s, S)存储策略。

9.10 某商店经销一种电子产品，每台进货价为 4 000 元，单位存储费为 60 元，如果缺货，缺货费为 4 300 元，每次订购费为 5 000 元，根据资料分析，该产品销售量服从区间[75, 100]内的均匀分布，即

$$\phi(r)=\begin{cases}\dfrac{1}{25} & 75\leqslant r\leqslant 100\\ 0 & \text{其他}\end{cases}$$

期初库存为零，试确定(s, S)型存储策略中 s 及 S 的值。

9.11 某音乐舞厅与一家饮料厂签订了长期合同，订购瓶装饮料。合同规定交货时间不得拖延，据以往统计，在订货期间的需求量(瓶)服从 $N(1\,000, 250^2)$。若一次订货的手续费为 100 元，每年平均需求量为 12 000 瓶，一瓶饮料一个月的保管费为 0.012 5 元，为使不缺货的概率达到 95%，问订货点是多少？每次订货多少？(每年按 360 天计算)



实际操作训练

(航空机票超订票问题)某航空公司执行两地的飞行任务，已知飞机的有效载客量为150人。按民用航空管理有关规定：旅客因有事或误机，机票可免费改签一次，此外也可在飞机起飞前退票。航空公司为了避免由此发生的损失，采用超量订票的方法，即每班售出票数大于飞机载客数。但由此会发生持票登机旅客多于座位数的情况，在这种情况下，航空公司让超员旅客改乘其他航班，并给旅客机票价的20%作为补偿。现假设两地的机票价为1 500元，每位旅客有 0.04 的概率发生有事、误机或退票的情况，问航空公司多售出多少张票？使该公司的预期损失达到最小。(**提示：**买票旅客登机与否服从二项分布。)

北京大学出版社版权所有
禁止转载

高等院校物流专业创新规划教材

序号	书　名	书　号	编著者	定价	序号	书　名	书　号	编著者	定价
1	物流工程	7-301-15045-0	林丽华	30.00	41	物流系统优化建模与求解	7-301-22115-0	李向文	32.00
2	物流管理信息系统	7-301-16564-5	杜彦华	33.00	42	物流管理	7-301-22161-7	张佺举	49.00
3	现代物流学	7-301-16662-8	吴　健	42.00	43	运输组织学	7-301-22744-2	王小霞	30.00
4	物流英语	7-301-16807-3	阚功俭	28.00	44	物流金融	7-301-22699-5	李蔚田	39.00
5	采购管理与库存控制	7-301-16921-6	张　浩	30.00	45	物流系统集成技术	7-301-22800-5	杜彦华	40.00
6	物料学	7-301-17476-0	肖生苓	44.00	46	商品学	7-301-23067-1	王海刚	30.00
7	物流项目招投标管理	7-301-17615-3	孟祥茹	30.00	47	项目采购管理	7-301-23100-5	杨　丽	38.00
8	物流运筹学实用教程	7-301-17610-8	赵丽君	33.00	48	电子商务与现代物流	7-301-23356-6	吴　健	48.00
9	现代物流基础	7-301-17611-5	王　侃	37.00	49	国际海上运输	7-301-23486-0	张良卫	45.00
10	现代物流管理学	7-301-17672-6	丁小龙	42.00	50	物流配送中心规划与设计	7-301-23847-9	孔继利	49.00
11	供应链库存管理与控制	7-301-17929-1	王道平	28.00	51	运输组织学	7-301-23885-1	孟祥茹	48.00
12	物流信息系统	7-301-18500-1	修桂华	32.00	52	物流案例分析	7-301-24757-0	吴　群	29.00
13	城市物流	7-301-18523-0	张　潜	24.00	53	现代物流管理	7-301-24627-6	王道平	36.00
14	营销物流管理	7-301-18658-9	李学工	45.00	54	配送管理	7-301-24848-5	傅莉萍	48.00
15	物流信息技术概论	7-301-18670-1	张　磊	28.00	55	物流管理信息系统	7-301-24940-6	傅莉萍	40.00
16	物流配送中心运作管理	7-301-18671-8	陈　虎	40.00	56	采购管理	7-301-25207-9	傅莉萍	46.00
17	物流工程与管理	7-301-18960-3	高举红	39.00	57	现代物流管理概论	7-301-25364-9	赵跃华	43.00
18	商品检验与质量认证	7-301-10563-4	陈红丽	32.00	58	物联网基础与应用	7-301-25395-3	杨　扬	36.00
19	供应链管理	7-301-19734-9	刘永胜	49.00	59	仓储管理	7-301-25760-9	赵小柠	40.00
20	逆向物流	7-301-19809-4	甘卫华	33.00	60	采购供应管理	7-301-26924-4	沈小静	35.00
21	集装箱运输实务	7-301-16644-4	孙家庆	34.00	61	供应链管理	7-301-27144-5	陈建岭	45.00
22	供应链设计理论与方法	7-301-20018-6	王道平	32.00	62	物流质量管理	7-301-27068-4	钮建伟	42.00
23	物流管理概论	7-301-20095-7	李传荣	44.00	63	物流成本管理	7-301-28606-7	张　远	36.00
24	供应链管理	7-301-20094-0	高举红	38.00	64	供应链管理(第2版)	7-301-27313-5	曹翠珍	49.00
25	企业物流管理	7-301-20818-2	孔继利	45.00	65	现代物流信息技术(第2版)	7-301-23848-6	王道平	35.00
26	物流项目管理	7-301-20851-9	王道平	30.00	66	物流信息管理(第2版)	7-301-25632-9	王汉新	49.00
27	供应链管理	7-301-20901-1	王道平	35.00	67	物流项目管理(第2版)	7-301-26219-1	周晓晔	40.00
28	物流学概论	7-301-21098-7	李　创	44.00	68	物流运作管理(第2版)	7-301-26271-9	董千里	38.00
29	航空物流管理	7-301-21118-2	刘元洪	32.00	69	物流技术装备(第2版)	7-301-27423-1	于　英	49.00
30	物流管理实验教程	7-301-21094-9	李晓龙	25.00	70	物流运筹学(第2版)	7-301-28110-9	郝　海	45.00
31	物流系统仿真案例	7-301-21072-7	赵　宁	25.00	71	交通运输工程学(第2版)	7-301-28602-9	于　英	48.00
32	物流与供应链金融	7-301-21135-9	李向文	30.00	72	第三方物流(第2版)	7-301-28811-5	张旭辉	38.00
33	物流信息系统	7-301-20989-9	王道平	28.00	73	国际物流管理(第2版)	7-301-28927-3	柴庆春	49.00
34	现代企业物流管理实用教程	7-301-17612-2	乔志强	40.00	74	现代仓储管理与实务(第2版)	7-301-28709-5	周兴建	48.00
35	出入境商品质量检验与管理	7-301-28653-1	陈　静	32.00	75	物流配送路径优化与物流跟踪实训	7-301-28763-7	周晓光	42.00
36	物流项目管理	7-301-21676-7	张旭辉	38.00	76	智能快递柜管理系统实训	7-301-28815-3	杨萌柯	39.00
37	智能物流	7-301-22036-8	李蔚田	45.00	77	物流信息技术实训	7-301-28807-8	周晓光	38.00
38	物流决策技术	7-301-21965-2	王道平	38.00	78	电子商务网站实训	7-301-28831-3	邢　颖	45.00
39	新物流概论	7-301-22114-3	李向文	34.00	79	电子商务与快递物流	7-301-28980-8	杨萌柯	49.00
40	库存管理	7-301-22389-5	张旭凤	25.00					

北京大学出版社版权所有 禁止转载

如您需要浏览更多专业教材，请扫下面的二维码，关注北京大学出版社第六事业部官方微信(微信号：pup6book)，随时查询专业教材、浏览教材目录、内容简介等信息，并可在线申请纸质样书用于教学。



感谢您使用我们的教材，欢迎您随时与我们联系，我们将及时做好全方位的服务。联系方式：010-62750667，63940984@qq.com，pup_6@163.com，lihu80@163.com，欢迎来电来信。客户服务 QQ 号：1292552107，欢迎随时咨询。